Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

www.diw.de

SOEP The German Socio-Economic Panel Study

SOEPpapers
on Multidisciplinary Panel Data Research

273

Jörg-Peter Schräpler

Benford's Law as an instrument for fraud detection in surveys using the data of the Socio-Economic Panel (SOEP)

Berlin, February 2010

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | urahmann@diw.de

Benford's Law as an instrument for fraud detection in surveys using the data of the Socio-Economic Panel (SOEP)*

Jörg-Peter Schräpler†

Abstract

This paper focuses on fraud detection in surveys using Socio-Economic Panel (SOEP) data as an example for testing newly methods proposed here. A statistical theorem referred to as Benford's Law states that in many sets of numerical data, the significant digits are not uniformly distributed, as one might expect, but rather adhere to a certain logarithmic probability function. To detect fraud we derive several requirements that should, according to this law, be fulfilled in the case of survey data. We show that in several SOEP subsamples, Benford's Law holds for the available continuous data. For this analysis, we have developed a measure that reflects the plausibility of the digit distribution in interviewer clusters. We are able to demonstrate that several interviews that were known to have been fabricated and therefore deleted in the original user data set can be detected using this method. Furthermore, in one subsample, we use this method to identify a case of an interviewer falsifying ten interviews who had not been detected previously by the fieldwork organization. In the last section of our paper, we try to explain the deviation from Benford's distribution empirically, and show that several factors can influence the test statistic used. To avoid misinterpretations and false conclusions, it is important to take these factors into account when Benford's Law is applied to survey data.

Keywords: Falsification, data quality, Benford's Law, SOEP
JEL classification: C69, C81, C83

*This paper is part of a research project at the DIW Berlin. I would like to thank Gert G. Wagner (DIW Berlin, Berlin University of Technology, TUB) and Peter Krause (DIW Berlin) for their assistance and valuable suggestions as well as Uli Pötter (DJI, München) and Christian Dudel (Ruhr-University Bochum) for helpful comments. The usual disclaimer applies.

†The author is a scientific officer of the Statistical Office of North Rhine Westphalia (IT.NRW, Düsseldorf), adjunct Professor at the Ruhr-University Bochum and a permanent visiting fellow of the DIW Berlin.

1 Introduction

In any survey in which the data are collected by personal interviews, there is a risk that interviewers may cheat, or that some may fabricate data. We can distinguish several forms of cheating.

Firstly, the most blatant form is when an interviewer fabricates all 'responses' in an entire questionnaire. The US Bureau of the Census refers to this practice as 'falsification' or 'fabrication'. Sometimes this practice is also unofficially called 'curbstoning', thus named because a census taker "stands at the curb" and guesses the number of residents in a building or house without ever entering. Interviewers who do this are called curbstoners[1] (Moore and Marquis 1996).

Secondly, a more subtle form is when an interviewer asks some questions in an interview and fabricates the responses to others.

A third form of cheating is when an interviewer knowingly deviates from prescribed interviewing procedures, such as conducting an interview with someone who is easily reachable and willing to participate in the place of the appropriate person.

Falsification might also include the acceptance of proxy information when self-response is required and the unauthorized use of the telephone when a personal visit is required.

In this paper we deal only with the first form of cheating: the fabrication of an entire interview. We focus on fabricated data in the German Socio-Economic Panel (SOEP) and apply an unconventional benchmark by the name of Benford's Law, which has already been used by several accountants to uncover fraud. Benford's Law is now also being used by several researchers in the social sciences to detect fabricated survey data (Diekmann 2002; Swanson et al. 2003; Schräpler 2004; Schräpler and Wagner 2005; Schäfer et al. 2005; Bredl et al. 2008) and frauds in regression coefficients in economics and the social sciences (Tödter 2009; Diekmann 2007). In our paper we try to give some explanatory notes for this logistic distribution and explore the effectiveness of this procedure in the case of survey data.

2 Previous results on cheating behavior

In comparison with other methodological topics, the literature contains only a few studies dealing with cheating by interviewers. Crespi (1945) described several factors that may contribute to cheating behavior. He distinguished between factors relating to questionnaire characteristics (design and length, difficult and antagonistic questions), administrative demoralizers (inadequate interviewer remuneration and training), as well as external factors (bad weather, bad neighborhoods, etc.). He proposed a dual strategy of eliminating demoralizers and using a verification method to deter cheating. Some more recent studies have referred to these verification methods and dealt with optimal designs of quality control samples to detect interviewer cheating (Biemer and Stokes 1989) and the evaluation of quality control procedures for interviewers (Stokes and Jones 1989).

Because of the lack of factual information concerning the nature of interviewer falsification, in 1982 the US Census Bureau implemented an "Interviewer Falsification Study" (Schreiner, Pennie, and Newbrough 1988). In this study data were compiled from fifteen surveys conducted by twelve US Census Bureau regional offices over a five-year period. They found 205 cases of confirmed falsification. Most of these (74%) were detected through re-interviews, and the majority (79%) were determined to have been fabricated interviews. Their results provide evidence that the shorter the length of service, the more likely an interviewer is to falsify data (Schreiner, Pennie, and Newbrough 1988). Furthermore, when new interviewers falsify data, they usually do so for a relatively high proportion of their assignments, and they tend to fabricate entire interviews. Interviewers with five or more years of experience usually falsify a smaller proportion of their assignments and tend to classify eligible units as ineligible (Hood and Bushery 1997).

[1] Curbstoning is a term that originated with 18th-century census-taking. This term was coined when it was discovered that some interviewers simply filled out interview forms without even contacting a respondent.

Other studies have dealt with the 'quality' of faked interviews and the impact of fabricated data on substantive analysis. Reuband (1990) showed that students are able to reproduce data in fictive interviews using available demographic variables on real respondents.

Schnell (1991) performed a study in which he substituted 220 real interviews from the German General Social Survey (ALLBUS 1988, N = 3,052) with fictive interviews fabricated by sociology students and their fellow students at the same university. He analyzed the quality of the fabricated data and the robustness of substantive empirical results by comparing the German General Social Survey with the substituted false data. His main result was that univariate statistics like proportions, means, and variances are relatively robust against typical amounts of fabricated data in surveys (less than 5%). Nevertheless, he also found some minor effects on multivariate statistics such as multiple regressions. Moreover, using simulations, he showed that higher proportions of fabricated data in surveys have a serious impact on multivariate statistics and data quality.

In the ALLBUS 1994, the ADM design was replaced with a new sampling design that offers the opportunity to systematically check that the interviews (N = 3,505) were performed correctly. The interviewers were given the names and address of the respondents directly. In six percent of the cases, irregularities were detected; half of them were falsified by the interviewers (Koch 1995). These fabricated data (n = 45) were found after the routine monitoring by the data collection institute via the postcard method, which detected fifteen falsified interviews in this survey. Another finding was that interviewers who cheat are mainly younger people with higher levels of education (Abitur) and with a relatively high workload (number of interviews). The SOEP is aware of the interviewer characteristics of those who cheat and was therefore able to compare them with the characteristics found in the ALLBUS (see Schräpler/Wagner 2005).

A rare debacle caused by falsified interviews is referred to by Diekmann (2002). In the German city of Rostock, a traffic study about drivers was carried out by means of 600 face-to-face interviews. Eighty cases were later re-contacted for another study, which showed that sixteen of the former interviews were completely or partly fabricated by the interviewer. If we extrapolate this to the whole sample, that amounts to a share of 20% fakes.

3 Benford's Law

Besides the 'conventional' tests for stability and consistency, an unconventional benchmark by the name of Benford's Law has recently been used by several accountants to detect frauds. Social researchers have also proposed using this method for survey data (Diekmann, 2002). In this and the following chapter, we will test whether Benford's Law can be used as an instrument for quality control and fraud detection in surveys.

Benford's Law is an empirical 'law' which states that in many tables of numerical data, the significant digits are not uniformly distributed, as one might expect, but rather adhere to a certain logarithmic probability distribution (Hill 1996b). According to Hill (1999), in 1881, the astronomer Newcomb (Newcomb 1881) explained that his discovery of the significant digit law was sparked by an observation that the pages of a book of logarithms were dirtiest in the beginning and progressively cleaner throughout. Nevertheless, the law is named after Dr. Frank Benford, a physicist who had made the same observation in 1938, when he embarked on a mathematical analysis of 20,229 sets of numbers, including such wildly disparate categories as the areas of rivers, baseball statistics, numbers in magazine articles and street addresses (see table 1, Benford 1938).

He found that all these seemingly unrelated sets of numbers followed the same first-digit probability pattern.[2] In most cases the number 1 turned up as the first digit about 30 percent of the

[2] Nevertheless Benford made no attempt to assess how good the fit was. On closer inspection of table 1, we can see that for some of these data sets, the digit frequency is not even a monotonically decreasing function of digit magnitude for higher valued digits. Using χ^2-tests Scott/Fasli (2001, p.5) show that only three of these data sets (D,F and R) have remarkably close fits, eight (A,G,I,M,O,P,Q and T) satisfy the standard 5% significant criterion, the remaining nine sets of data cannot be regarded as conforming to Benford's Law.

Table 1: *The distribution of leading digits in Benford's data sets in percentages (Benford 1938)*

Group	Title	1	2	3	4	5	6	7	8	9	Count
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3,259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1,389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P.Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Weight	27.7	25.3	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1,800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	n^{-1}, $\sqrt{n}$,...	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5,000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	Gigest	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1,458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1,165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2, \ldots, n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
	Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1,011
	Predicted	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6	

time, more often than any other. Benford derived a formula to predict the frequency of numbers found in many categories of statistics. The leading significant (non-zero) digit obeys the law

$$Prob(\text{first significant digit} = d) = log_{10}\left(1 + \frac{1}{d}\right), \qquad d = 1, 2, \ldots, 9$$

Hence, a number chosen at random has leading significant digit $d = 1$ with probability 0.301, a leading digit $d = 2$ with probability 0.176 and so on monotonically down to probability 0.046 for leading digit $d = 9$. The general law for second and higher significant digits and their joint distribution is (Hill 1996a, 1999):

$$Prob(D_1 = d_1, \ldots, D_k = d_k) = log_{10}\left[1 + \left(\sum_{i=1}^{k} d_i \times 10^{k-i}\right)^{-1}\right] \tag{1}$$

where $d_1 \in \{1, 2, \ldots, 9\}$ and $d_j \in \{0, 1, 2, \ldots, 9\}, j = 2, \ldots, k$. Therefore the joint probability $Prob(D_1 = 1, D_2 = 5, D_3 = 2) = log_{10}(1 + (152)^{-1}) \approx 0.0028$.

From equation 1 follows that the significant digits are dependent and not independent. In the appendix, table 16 shows the joint distribution for the first two digits. It can easily be seen that the joint probability that the second digit is 3, given that the first digit is 1, is $P(D_1 = 1, D_2 = 3) \approx 0.0322$, but $P(D_1 = 1) \cdot P(D_2 = 3) \approx 0.0314$.

This interdependence among significant digits decreases rapidly as the distances between the digits increases. The table below table 16 shows the distribution of the first to the fourth significant digits. We can recognize that the distribution of the nth significant digit approaches the uniform distribution on $0, 1, \ldots, 9$ exponentially fast as $n \to \infty$ (c.f. Hill 1995, p.355).

For many years, this law was considered little more than a numerical curiosity, but practical implications began to emerge in the 1960s (Scott/Fasli 2001). It was recognized that the suggestion that almost 1/3 of the numbers processed began with the digit '1' could have implications for the design of efficient computers (Hamming 1970; Knuth 1981). In recent years Benford's Law has been used successfully to detect fraudulent financial data (Nigrini 1999).

Despite this rather slender empirical support (Scott/Fasli 2001), there is disagreement about whether this law is a necessary mathematical truth or a contingent property of nature.

3.1 Explanations of Benford's Law

3.1.1 Scale invariant Theorem

The literature contains several theoretical papers that have attempted to explain why Benford's Law is true. The first step towards explaining this relationship was taken in 1961 by the mathematician Roger Pinkham (Pinkham 1961). He argued that if there is a law of digit frequencies, it should be universal and 'scale-invariant.' This means that if we multiply all our numbers by an arbitrary constant, then the distribution of first-digit frequencies should remain unchanged. Pinkham provided the proof that if a law of digit frequencies is invariant under changes of scale (e.g., dollars to euros) then it has to be Benford's Law. Furthermore, Hill (1995) was able to show that scale invariance implies base invariance, but not conversely.[3] Nevertheless, this explanation makes no contribution to answering the question as to whether real data should conform to the logarithmic law.

3.1.2 Multiplying a lot of numbers together

Another approach is based on the notion of producing a number by multiplying a lot of numbers together.[4] Boyle (1994) showed that the logarithmic distribution is the limiting distribution of leading digits when random variables are repeatedly multiplied, divided, or raised to integers powers. Scott/Falsi (2001) were able to show in their simulations that there is indeed convergence towards the logarithmic distribution in all checked cases, and that for some distributions this convergence is rapid.[5]

3.1.3 The random-samples-from-random-distribution theorem by Hill (1995)

A plausible theoretical explanation for the appearance of this logarithmic distribution is the *random-samples-from-random-distribution theorem* by the mathematician Hill (1995). He showed "that if probability distributions are selected at random, and random samples are then taken from each of these distributions in any way so that the overall process is scale (or base) neutral, then the significant digit frequency of the combined sample will converge to the logarithmic distribution." (Hill 1995, p. 360). If Hill's theorem is correct, this means that the digits derived from a random mix of different sources, from census data to stock market prices, should follow Benford's Law. The mixture of data may be the key.[6] It is not a requirement that the individual realizations of a random variable have to be scale- or base-invariant. However, it is necessary that the sampling process on average does not favor one scale over another (Hill 1995, p.361). This theorem may be important in helping us to answer the question as to whether Benford's Law is feasible for survey data as survey data contain different variables with different distributions.

[3]However Lolbert (2008) attempted to show that there exists no probability measure that would obey Benford's law for any base, but if the set of possible bases does not exceed a given upper limit, most real-life distributions obey, or can be transformed to obey Benford's law.

[4]This approach is appealing because of its similarity to the Central Limit Theorem. The notion that Benford's Law might embody a similar general rule for the production of a number of random variables is very pleasing (Scott/Falsi 2001, p.4).

[5]They showed that two factors influence the rate of convergence: the variance of the mantissa of the random variate and the deviation of the random variate's distribution from Benford's Law.

[6]In table 1 Benford computed the average values for each of the digit frequencies. They look very close to the predicted values. However, it seems that he simply computed the average percentages for each digit and taken no account of the different sample sizes of the data sets.

3.2 Empirical evidence

There is evidence that many classes of true data sets follow Benford's Law. It has been found to apply to many sets of financial data, including income tax and stock exchange data, corporate disbursements and sales figures, demographic and scientific data (e.g., Nigrini 1999), as well as numbers gleaned from newspaper articles (Benford 1938; Hill 1999). In the case of non-random sequences Luque/Lacasa (2009) showed that Benford's Law describes with astonishing precision the statistical distribution of leading digits in the prime number sequence. Stock prices may seem to be a single distribution but their value actually stems from many measurements (salaries, the cost of raw material and labor) and so it is expected that they will follow Benford's Law in the long run. A recent study about whether tax returns in Germany follow Benford's Law showed that not all but the majority do conform to the logarithmic distribution[7] (Posch 2003).

In the case of stock market companies, which represent all stages of growth, Nigrini gave an additional intuitive explanation. We can consider a growing company with a market value of 100 million euros. For the value to reach 200 million euros, the company must double its value. For it to increase from 200 million euros to 300 million euros it must increase only 50%, and for it to increase from 900 million euros to 1000 million euros it must increase by just 11%. Moreover, for it to increase from 1,000 million euros to 2,000 million euros it must again double. Hence a growing company spends longer with a '1' as the first digit of its market capitalization than it does with any other number. The persistence of a 1 as a first digit will occur with any phenomenon that has a constant or erratic growth rate (Nigrini 1999).

On the other hand, truly random numbers do not confirm to Benford's Law because the proportion of leading digits in such numbers are, by definition, equal. Those data sets most likely to follow Benford's Law have numbers that do not contain a built-in maximum and describe the sizes of similar phenomena (Nigrini 1999). Assigned numbers, such as social security numbers or bank accounts, will not conform to it. Furthermore, deviations from the law's prediction can be caused by merely rounding numbers up and down. Moreover, the sample of numbers should be large enough to give the predicted proportions a chance to assert themselves (Pinkham 1961), and the sets of numbers should essentially be subsets of a larger series and not just huge chunks of that series.

Recently Benford's Law has been used to determine the normal level of number duplication in data sets, which in turn makes it possible to identify abnormal digit and number occurrence. Accountants and auditors have begun to apply Benford's law to corporate accounting to discover number pattern anomalies and frauds. Nigrini found that true tax data have a close fit to Benford, and there is substantial evidence that in most fabricated tax data the significant digits are not close to Benford. Usually the falsified data reveal conspicuous patterns and do not follow the expected distribution. Nigrini used a goodness-of-fit-to-Benford test and successfully identified fraudulent financial data.

3.3 Recent empirical explanations

The simulation results of Scott/Fasli (2001) Scott and Fasli (2001) pointed out that "the situation is thus such that, if it should turn out to be the case that Benford's Law is valid, then there are several alternative mathematical explanations of why this should be so. On the other hand, none of them imply that the logarithmic law is necessary true." (Scott/Fasli 2001, p.4). In their experimental study, they first tried to find 'natural data sets' that conform to Benford's Law. They investigated 230 data sets, all of which can be accessed on the web. In total, over a half a million numbers were examined. They found that only 12.6% (29 of 230) satisfied the 5% significance criterion for conformity to Benford's Law. However, they found a significant number

[7] Posch (2003) investigated a data base with tax returns ($N = 21$ mill. records) from the year 2001 from a finance office in NRW. His results show that foreign earnings, earnings from independent personal services, from leasing, and from capital assets closely adhere to Benford's Law. Gross earnings do not follow the logarithmic distribution.

of real data sets that definitely do not conform to the law but have leading digit distributions that are broadly similar. In particular, leading digit frequency proved to be a monotonically decreasing function of digit value.

In mathematically generated data sets they investigated, in a second step, recurrent products and products of random variates.[8] The main results are (Scott/Fasli 2001):

- Multiplying the current number by a constant: in the majority of cases the resulting distribution is very close to Benford's Law. The exceptions arise when the multiplier is an exact integral power or root of 10 because multiplying by 10 does not change the leading digit.
- Multiplying the current number by a uniformly distributed random variate: such sequences also converge to the logarithmic distribution except in those cases where the mean is an integral power or root of 10 and the standard deviation is small.
- Each number in the data set is the product of several random variables: the results show convergence toward the logarithmic distribution in all cases. Two factors influence the rate of convergence: the variance of the mantissa of the random variate and the deviation of the random variate's distribution form Benford's Law.

Benford's Law and the lognormal distribution These results support the theoretical models that are based on recurrent multiplication and on the assumption that each item is the product of several random variates. The latter is the equivalent to adding their logarithms. Because the sum of independent random variates tends to a normal distribution as the number increases (central limit theorem), the logarithm of the product of random variates should also tend to a normal distribution. Therefore, there is a connection between Benford's Law and the lognormal distribution. Scott/Falsi showed that conformity to Benford's Law is a function of the shape parameter σ and independent of scale parameter (median) because Benford's Law is scale-invariant (Hill 1996a). Very good fits appear if the shape parameters of the lognormal distributions exceed the value 1.2. Scott/Falsi concluded from this finding that data, the distribution of which conforms to a lognormal distribution and the shape of which exceeds 1.2, should give rise to leading digit distributions satisfying the logarithmic law. This is the case if:

1. the data set has only positive values
2. the data set has a unimodal distribution whose modal is not zero
3. the data set has a positive skewed distribution in which the median is no more than half of the mean.

The latter ensures that the shape parameter of the lognormal distribution will exceed 1.2. From their empirical results Scott/Falsi drew their fundamental conclusion that "Benford's Law is not a necessary mathematical truth or a deep mystical property of our universe. It is a straightforward consequence of the way in which we quantify our observations of that universe. Measurements that cannot meaningfully take values less than zero give rise to Benford's Law. Not all of them do. If the range of measurement is such that zero falls well outside the range of practical consideration, then the leading digits will not conform to the law. But many of the quantities that we measure are necessarily positive and have ranges that include significant numbers of items close to zero. According to our explanation, it is these that give rise to Benford's Law." (Scott/Falsi 2001, p.17).

Therefore Scott/Fasli concluded, on the basis of their simulation results, that many real data sets conform to Benford's Law because their distribution follows a lognormal distribution with a

[8]There have been several other simulation studies. Engel/Leuenberger (2003) showed in their study that exponentially distributed random numbers obey Benford's law approximatively, i.e., within bounds of 0.03. Miller/Nigrini (2008) explained why so many data sets follow Benford's Law (or at least a close approximation to it). They showed that if we can consider the observed values of a system to be the product of many independent processes with reasonable densities, then the distribution of the digits of the resulting product will come close to Benford's Law.

shape greater than 1.2. They stated that a large number of naturally occurring quantities have these characteristics.

4 Using Benford's Law on survey data?

An interesting point for survey researchers is whether this logarithmic distribution, Benford's Law, can also be used to identify fabricated data in surveys. Hence, the main question is whether survey data follow Benford's Law. Unlike financial data, many variables in these databases are dichotomous or categorical (like gender, marital status, and occupation) or are assigned numbers like household numbers. In this case, these data certainly do not conform to Benford's Law. However there are often also variables which refer to other monetary or continuous values.

4.1 SOEP data and their confirmation to Benford's Law

For our analysis we use the Socio-Economic Panel Study (SOEP). The SOEP is a longitudinal representative survey containing socioeconomic information on private households in the Federal Republic of Germany (Wagner et al. 2007). It is similar to the US Panel Study of Income Dynamics (PSID). DIW Berlin (German Institute for Economic Research) manages the SOEP study. The first wave of data, collected in 1984 in the old Federal Republic of Germany, contains 5,921 households. The original sample was supplemented by a sample of East German residents (sample C) in 1990 (2,179 households) and a sample of immigrants in 1994-1995 (sample D, 522 households). Additional refreshment samples were added in 1998 (sample E, 1,056 households), 2000 (sample F, 6,052 households), and 2006 (sample H, 1,506 households).[9] All household members aged 16 and older are interviewed. For our analysis we use the first waves of the samples A/B, C, E, and F.

4.1.1 Requirements

The literature shows that the validity of Benford's Law depends on certain conditions. We try to summarize all necessary requirements that have to be fulfilled in order to detect fraudulent data in surveys with Benford. Some of these requirements are derived from simulation results (Scott/Fasli 2001), others are findings from practical applications (Nigrini 1999) or theoretical analyses (Hill 1995).

- The data set should not contain a built-in maximum because the frequency of these values will occur more often in the digit analysis and will cause biased results (Nigrini 1999).
- The data set should not contain assigned numbers such as social security numbers or bank accounts (Nigrini 1999).
- The data set should only have positive values with a unimodal distribution whose modal is not zero (Scott/Fasli 2001).
- The data set should have a positive skewed distribution in which the median is lower than the mean. Hence, the data set should contain more smaller than larger values.
- The data set should not emanate from statistical procedures like calculated means or variances that emanate from other data (Mochty 2002).
- The usefulness for survey data depends on the existence of continuous variables in the data set. Survey data that only contains categorical data does not meet the aforementioned conditions.

[9]SOEP also provides data about interviewer characteristics (see Schräpler and Wagner 2001).

A further requirement is a large enough sample size of the data set. The larger the sample size, the better the fit to Benford's distribution should be, as long as all of the above requirements are satisfied.

4.1.2 Description of the data used for Benford's Law

In the first step, we give a short description of the data we use. The selected data are restricted to variables with monetary values. Apart from monthly gross and net income, the data sets contain variables like gross amount of Christmas or vacation bonus, gross amount of monthly unemployment benefits or monthly subsistence allowance, gross amount of early retirement benefits, amount of taxes, as well as many other monetary variables. The amount of monetary variables increases over the waves. The first two waves in the years 1984 and 1985 contain about twenty variables and this increases to over thirty in subsequent waves. Sample C starts in 1990 with over 40 monetary variables and samples E and F contain about 60 variables in the year 2000.[10]

The monetary values are pooled over all variables for each selected wave. The distribution of these data sets is shown by a kernel density estimation method with an Epanechnikov function[11]. Other kernel functions like Gaussian or Parzen will result in quite similar distributions.

The figures 1, 2, and 3 show the estimated distributions for the waves used from samples A/B, C, E, and F. They contain the number of variables in the data set, the number of values (N), the mean and the standard deviation, as well as the median.

We can see that - except for figure 2 - all distributions tend to have a similar shape: the distributions are unimodal and the medians are always lower than the means and yield positive-skewed distributions. A unimodal positive-skewed distribution is one important requirement for the use of Benford's Law (Scott/Falsi 2001). The monetary data sets of sample C in figure 2 are unimodal but quite symmetric and not positive-skewed, the values for median and mean are quite close.

4.1.3 Wave-specific fit to Benford for several subsamples in SOEP

In the next step, we examine the overall goodness of fit for these datasets. The following figures 4 to 11 show the first digit and the first two digit distributions of the selected data in the first eight waves of samples A and B. The 95% confidence interval for the first digit distribution is calculated with (Nigrini 2000, p.43):

$$Upper = h_{b_d} + 1.96 \cdot \sqrt{h_{b_d} \cdot \frac{(1-h_{b_d})}{n}} + \frac{1}{2n} \tag{2}$$

$$Down = h_{b_d} - 1.96 \cdot \sqrt{h_{b_d} \cdot \frac{(1-h_{b_d})}{n}} - \frac{1}{2n} \tag{3}$$

where h_{b_d} is the expected proportion according to the logarithmic distribution and n indicates the sample size of all analyzed numbers. Unfortunately, the usefulness of these intervals is limited. Due to the fact that the sample size of the digits is larger than 20,000 in all waves, we get very close confidence intervals. Hence, even very small deviations from Benford's distribution are always statistically significant.

To examine the overall fit to Benford, the chi-square value has the disadvantage that it depends strongly on the sample size. One alternative is to use a measurement which relates to the worst possible fit. This is the case if all digits in one cluster have the most unlikely value, the digit 9. We define this goodness of fit (GFI) measurement with

[10] The labels of the variables used can be found in the appendix.

[11] Epanechnikov Kernel: the Epanechnikov kernel is this function: $(3/4)(1-u^2)$ for $-1 < u < 1$ and zero for u outside that range. Here $u = (x - x_i)/h$, where h is the window width and x_i are the values of the independent variable in the data, and x is the value of the scalar independent variable for which one seeks an estimate (Epanechnikov 1969). The window width used is calculated automatically using the statistics program STATA.

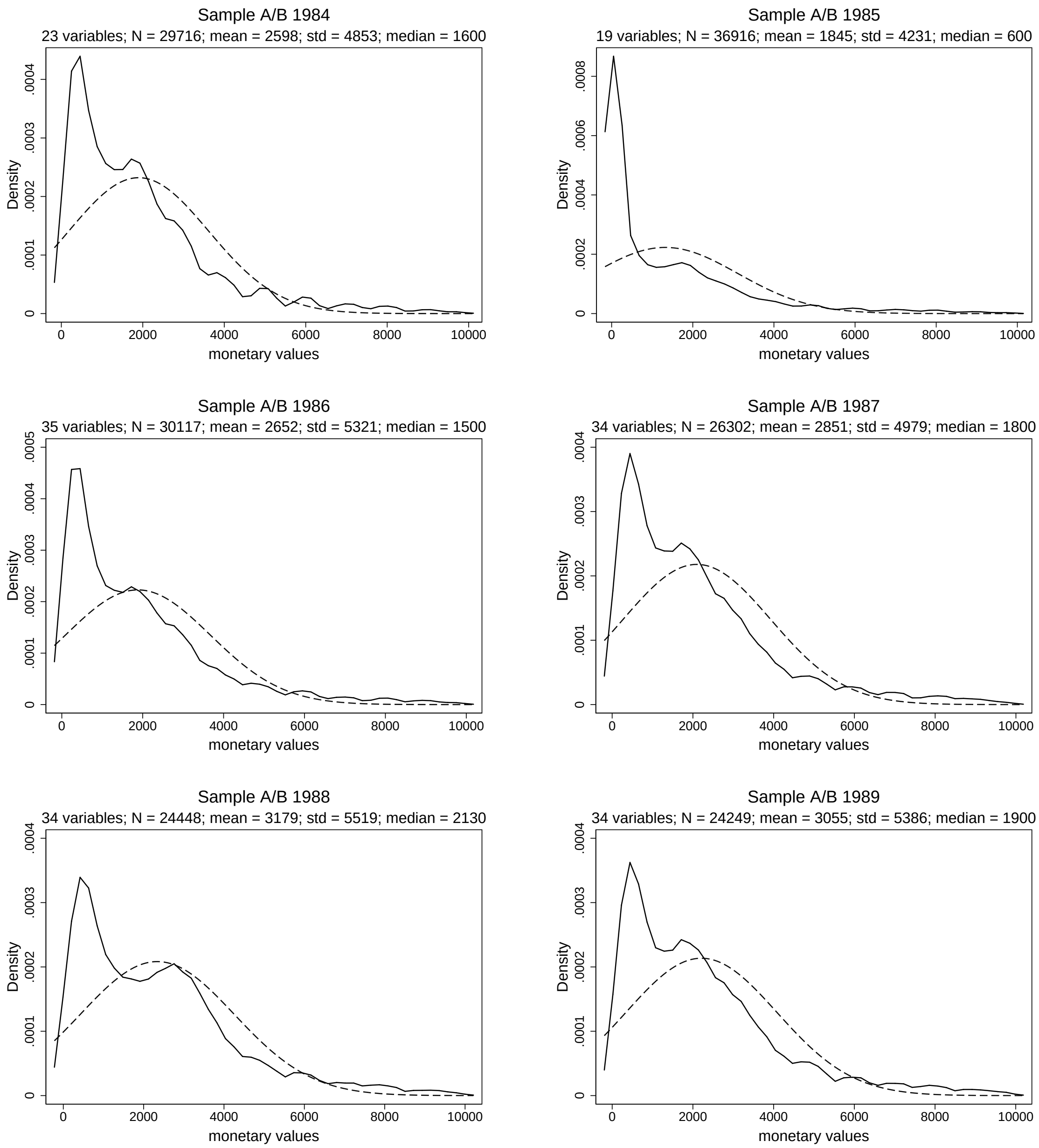


FIGURE 1: *Kernel density estimation for the distribution of the selected monetary data sets of sample A/B wave 1-6 in the SOEP (normal density function is dashed)*

$$GFI = 1 - \frac{\chi_i^2}{\chi_0^2} \quad \text{where} \quad i = 1, \ldots, n \tag{4}$$

the index i indicates the interviewer cluster and χ_0^2 is the chi-square value for the distribution with the worst fit to Benford's Law. The range is from 0 to 1, where the value 1 indicates an exact Benford distribution and values over 0.99 indicate a very close fit.[12]

[12]This measurement GFI is built in analogy to the well-known goodness of fit measurement GFI for LISREL

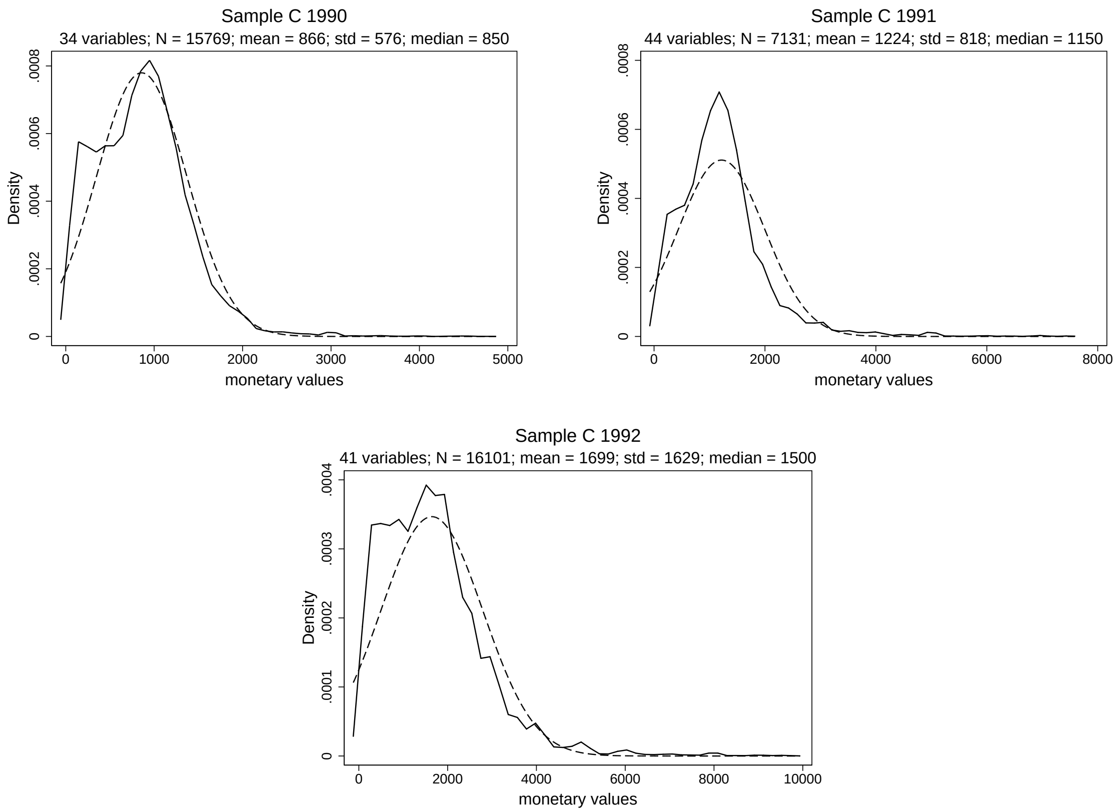


FIGURE 2: *Kernel density estimation for the distribution of the selected monetary data sets of sample C wave 1-3 in the SOEP (normal density function is dashed)*

Overall fit to Benford - the first eight waves of samples A/B If we take a look at the first four waves of samples A and B (figures 4-7), we find very similar distributions. At a first glance, we can see that the shape of the distributions are quite close to Benford. Nevertheless, on closer inspection, we can see that, in fact, the proportion of the first digit '1' is in line with Benford, but the proportions for the first digits '2' and '3' are significantly higher and for digits > 3 slightly lower than in the logarithmic distribution. However, the overall fit to Benford, measured with the GFI index seems to be very good: the values are close to 0.998 for waves 1-4. The next figures 8-11 show the first digit distributions for the following waves 5-8. The order of the frequencies of the first digits '1', '2', and '3' are still sustained but we can see that the proportion for digit '1' is distinctly lower than predicted and the proportion for the digits '2' and '3' increases over time. The GFI declines to a value of 0.996 and 0.995. One reason for the shift from the first digit '1' to higher digits might be the development of the monthly income. From waves 1 to 8 the average net income increases from DM 1,745 to DM 2,188 and the gross income from DM 2,552 to DM 3,199 in samples A and B. Many other monetary variables that are included in our descriptive analysis are related to this income variable.

The distributions for the first two digits in figures 4-11 show significantly higher proportions for numbers like $10, 20, 30, \ldots, 90$. We will see later that this finding is a result of the respondent's rounding behavior. Unlike many other collected data sets such as data from stock markets which contain relatively precise continuous monetary values, interview data are often rounded (cf.

models. However, there the fit value of an actual model refers to a value of the fit function for a model containing only a constant.

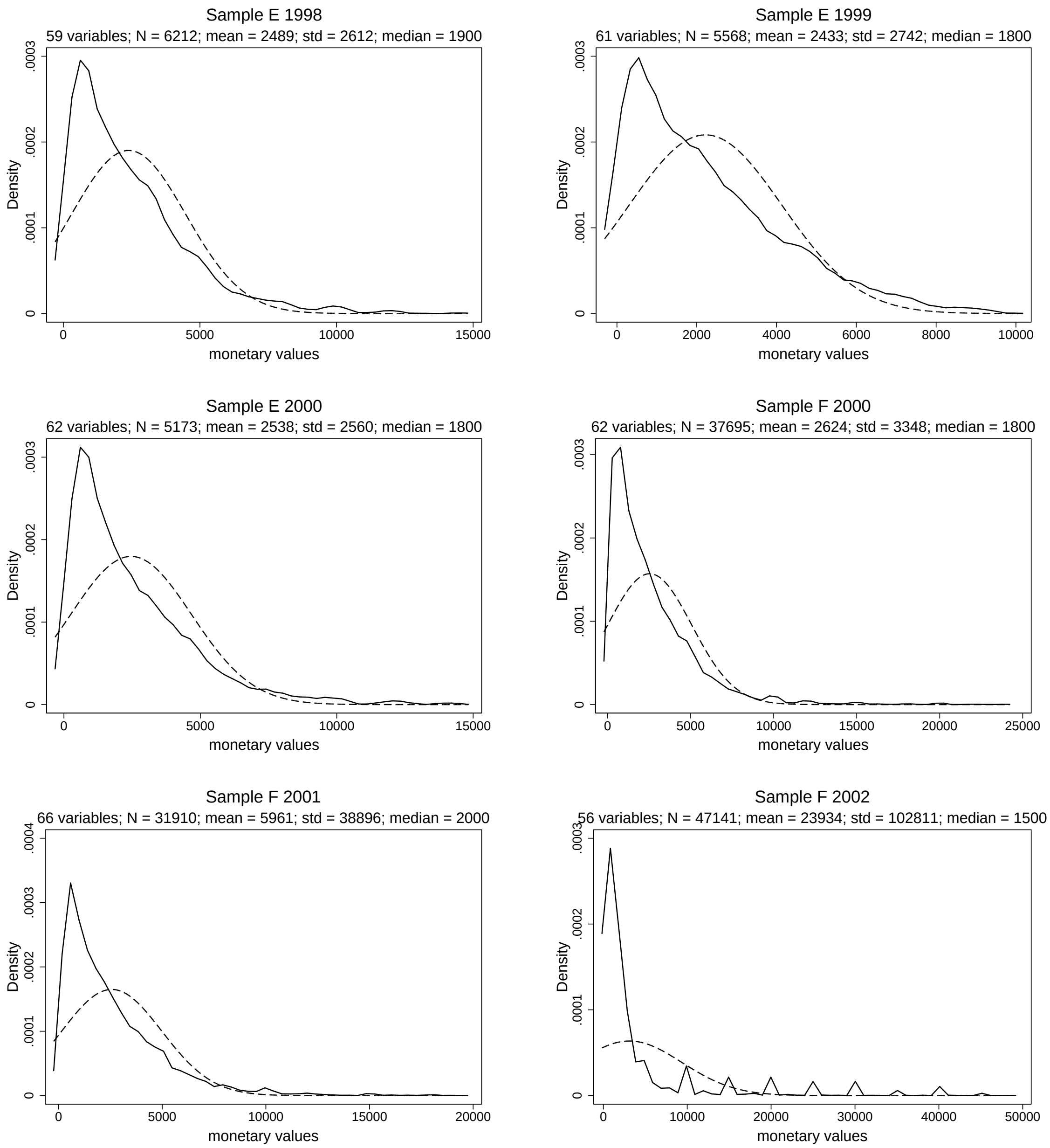


FIGURE 3: *Kernel density estimation for the distribution of the selected monetary data sets of sample E wave 1-3 and sample F wave 1-3 in the SOEP (normal density function is dashed)*

Schräpler 1999). The respondents often have cognitive problems recalling their exact gross income or other income-related variables. Therefore, the given values in surveys are more or less rough estimates and rounded after the first or second significant digit. The distributions of the first two digits give some information about this rounding behavior. We can see that the digit '30' has the highest peak, followed by the digits '20' and '10' in the first eight waves of the SOEP. Besides this rounding behavior, the figures show that the shape of these distributions have one characteristic in common with Benford's distribution: the proportion of smaller digits is higher than the proportion of larger digits.

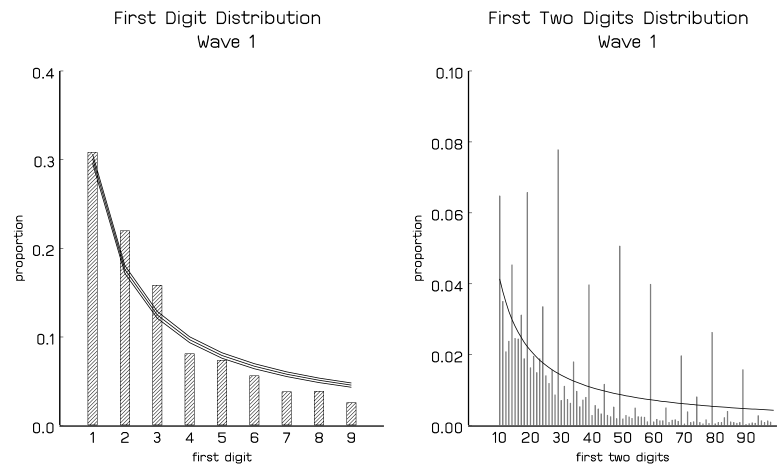


FIGURE 4: *Sample A/B, wave 1,* $\chi^2 = 1279$, *GFI = 0.9979 for first digit distribution,* $N = 29,712$

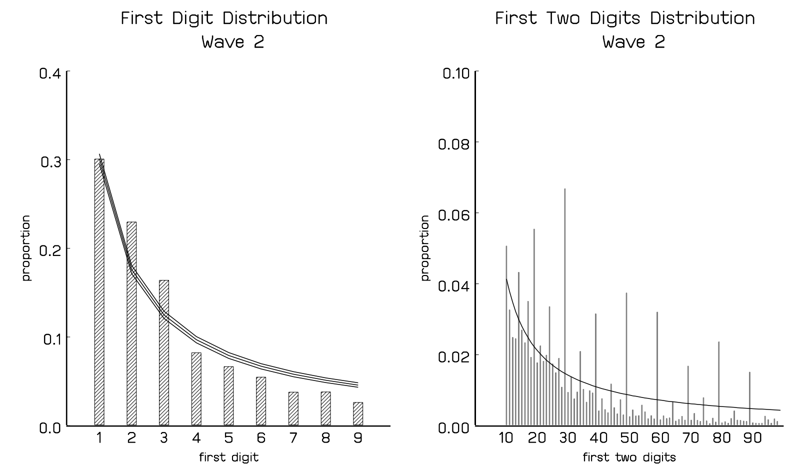


FIGURE 5: *Sample A/B, wave 2,* $\chi^2 = 1409$, *GFI = 0.997 for first digit distribution,* $N = 26,167$

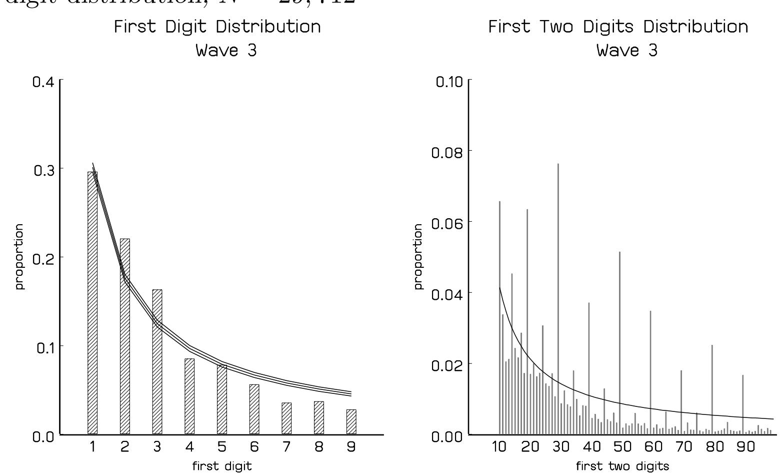


FIGURE 6: *Sample A/B, wave 3,* $\chi^2 = 1378$, *GFI = 0.9979 for first digit distribution,* $N = 30,113$

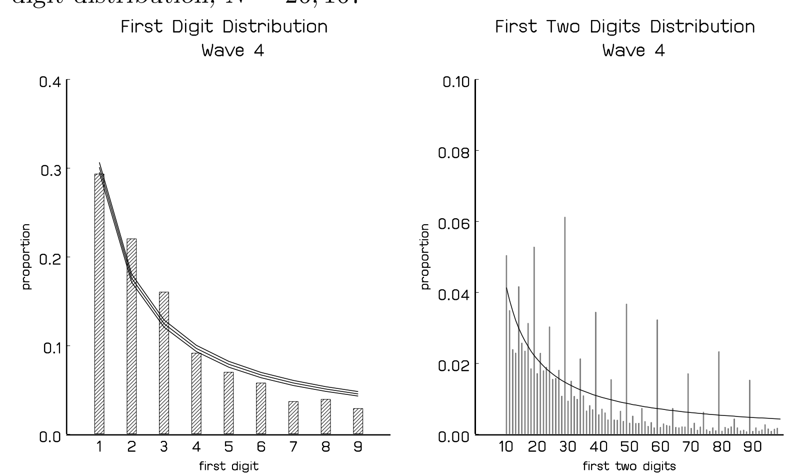


FIGURE 7: *Sample A/B, wave 4,* $\chi^2 = 1059$, *GFI = 0.998 for first digit distribution,* $N = 26,299$

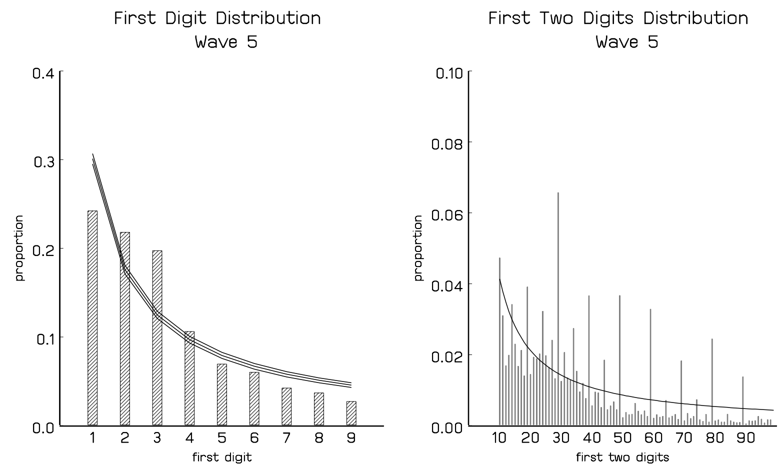


FIGURE 8: *Sample A/B, wave 5,* $\chi^2 = 2002.24$*, GFI = 0.996 for first digit distribution,* $N = 24,445$

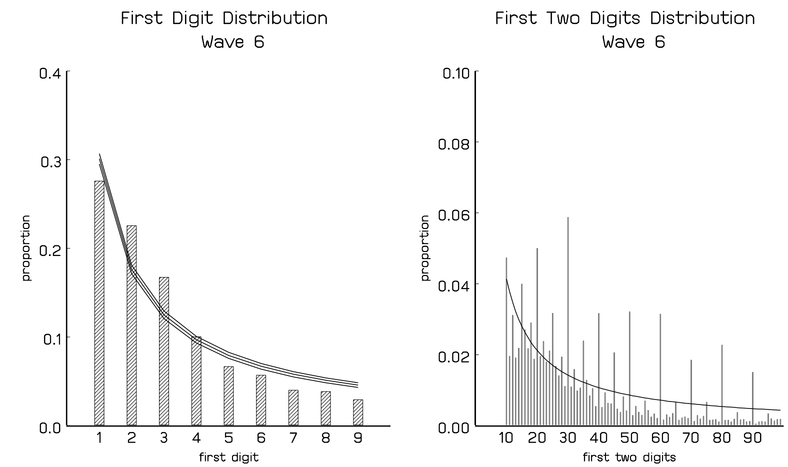


FIGURE 9: *Sample A/B, wave 6,* $\chi^2 = 1182.4$*, GFI = 0.998 for first digit distribution,* $N = 24,249$

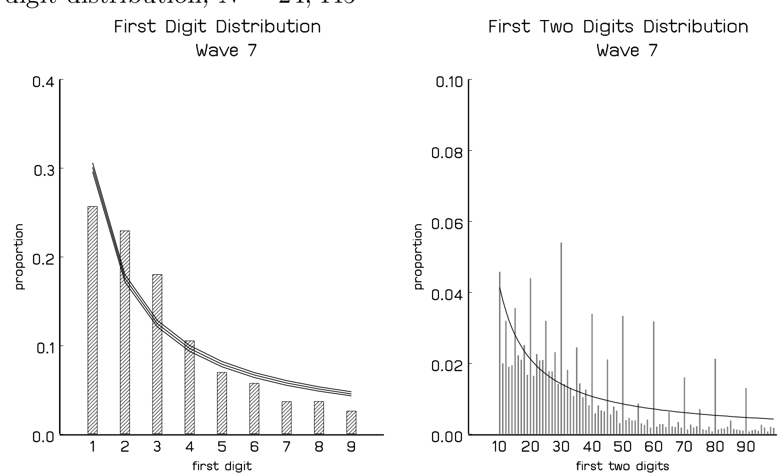


FIGURE 10: *Sample A/B, wave 7,* $\chi^2 = 2227.7$*, GFI = 0.997 for first digit distribution,* $N = 31803$

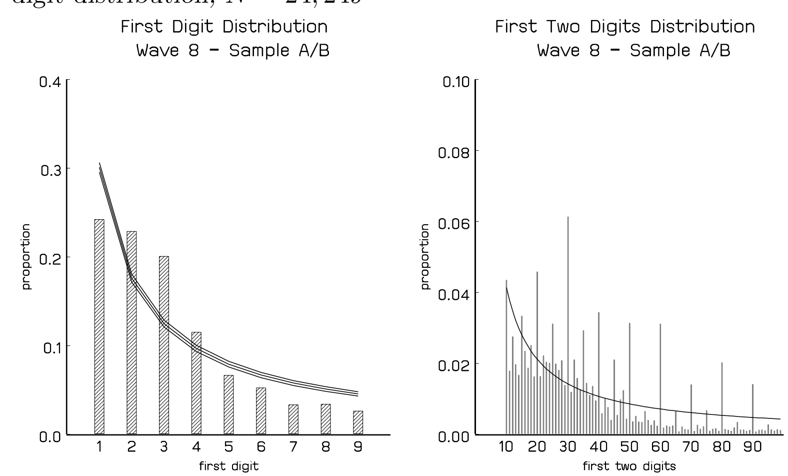


FIGURE 11: *Sample A/B, wave 8,* $\chi^2 = 2957.9$*, GFI = 0.995 for first digit dist.,* $N = 27,875$

Overall fit to Benford - the first three waves of sample C Figures 12-14 show the distribution in waves 1, 2, and 3 for the East German sample C (7, 8, and 9 for the SOEP). Obviously, we can recognize different patterns. The data were collected in the years 1990, 1991, and 1992, directly after German unification. In the years 1990 (wave 1), 1991 (wave 2), and 1992 (wave 3), the average gross income of East German residents increases from DM 811 to DM 1,555 and DM 2,089 and the net income from DM 667 to DM 1,172 and DM 1,508. We find too many monetary values with a first digit '1' in the data sets for waves 1 and 2, and in wave 1 also higher proportions than predicted for the digits '7', '8', and '9'. In figure 2 on page 11 we can see that the distributions in these data sets are symmetric, not positive-skewed, and that the variances are very small. In the years 1990 and 1991 the standard deviations are lower than the mean that result in variation coefficients ($std./mean$) lower than one. All other data sets analyzed have variation coefficients higher than one. The majority of the monetary values lie between DM 500 and 2,000 in the year 1990 and between DM 1,000 and 2,000 in the year 1991. Overall, this entails larger deviations from the logarithmic distribution. In the year 1992 (wave 3) we can observe a strong increase in monthly income and other monetary variables caused by the transformation and harmonization process. Therefore, the proportion of higher first digits like '2' to '5' increases, and the first digit distribution adheres more closely to Benford's distribution.

Overall fit to Benford - the first three waves in sample E Figures 15-17 show the digit distribution in wave 1 to 3 in sample E. The sample sizes are smaller than in samples A/B and sample C, which results in wider confidence intervalls. Although the overall shape is very similar to the logarithmic distribution, we find partly significant deviations from Benford. The proportions for the first digits '1', '7'-'9' are slightly lower and for the digits '2'-'6' slightly higher than the predicted proportions. Again, the first two digit distributions show the aforementioned characteristic rounding behavior.

Overall fit to Benford - the first three waves in sample F Figures 18-20 show the digit distribution in waves 1 to 3, sample F. The overall shape is quite similar to the logarithmic distribution in all three waves. Because of very high sample sizes ($N > 30,000$) we get close confidence intervals and significant differences from the predicted distribution for all digits.

Summary In this section, we have examined whether Benford's Law holds in the selected data sets to be sure that we can use the logarithmic distribution for detecting suspicious interviewer clusters. We expect that if the overall digit distribution in each wave does not closely adhere to Benford's distribution, we cannot continue to be sure that this will be the case in specific interviewer clusters.

The data sets used contain only continuous variables. Overall, our results show rather good fits to Benford's Law in the first waves of the subsamples A/B, E, and F of the SOEP. All these data sets are positive-skewed with the exception of subsample C, which shows a symmetric shape and large differences in respect to the anticipated logarithmic distribution. We therefore cannot expect that the use of Benford's Law will lead to satisfying results for sample C.

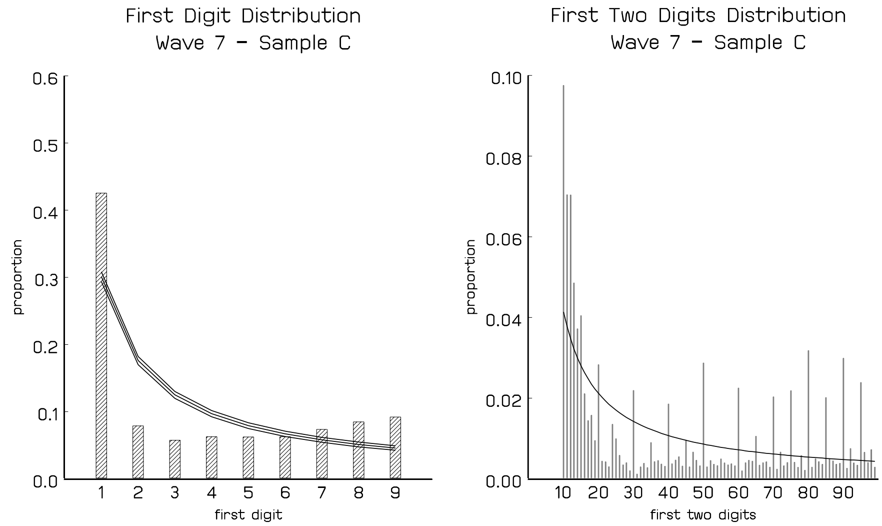


FIGURE 12: *First wave in sample C,* $\chi^2 = 3,626$*, GFI = 0.989 for first digit distribution,* $N = 15,769$*, only values with min. 2 digits*

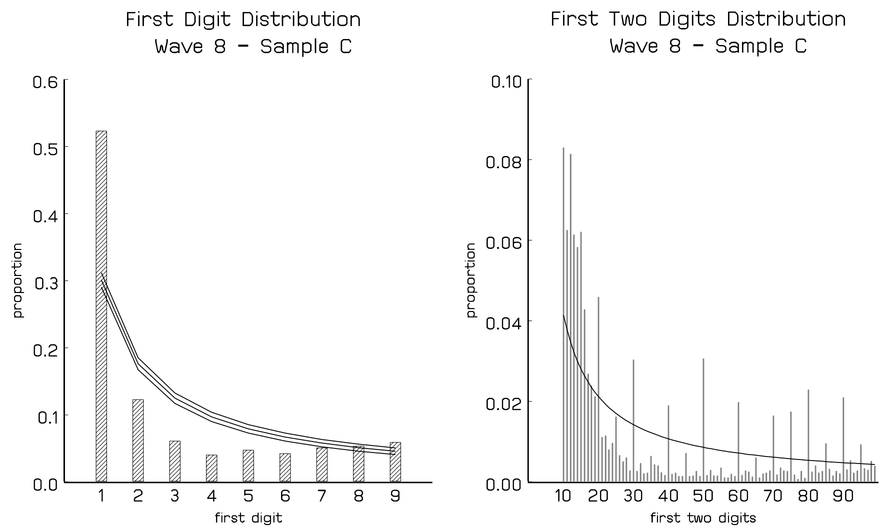


FIGURE 13: *Second wave in sample C,* $\chi^2 = 1941$*, GFI = 0.987 for first digit distribution,* $N = 7,126$

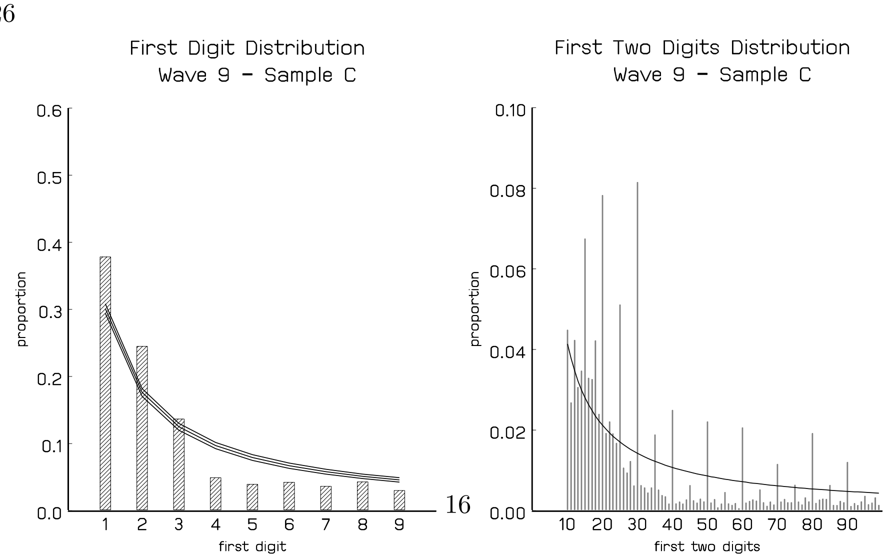


FIGURE 14: *Third wave in sample C,* $\chi^2 = 1861.3$*, GFI = 0.994 for first digit distribution,* $N = 16,101$

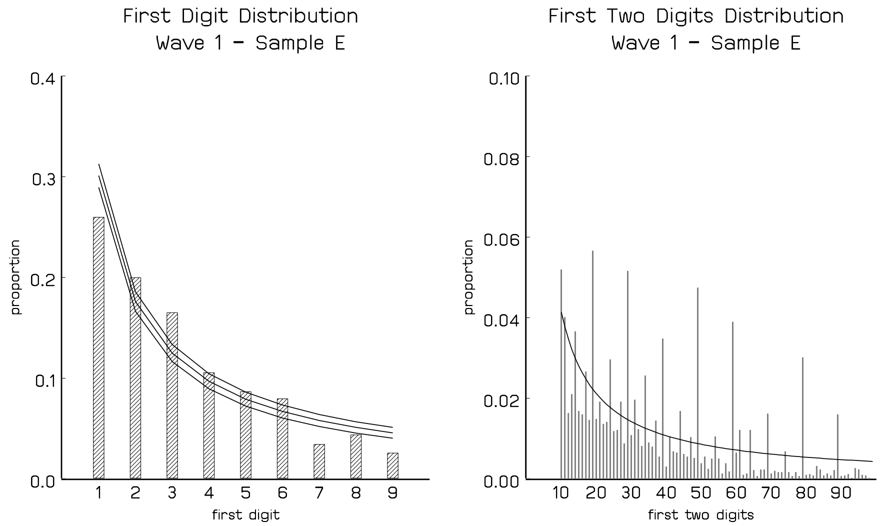


FIGURE 15: *Sample E, wave 1, $\chi^2 = 272$, GFI = 0.9979 for first digit distribution, $N = 6,212$, only values with min. 2 digits*

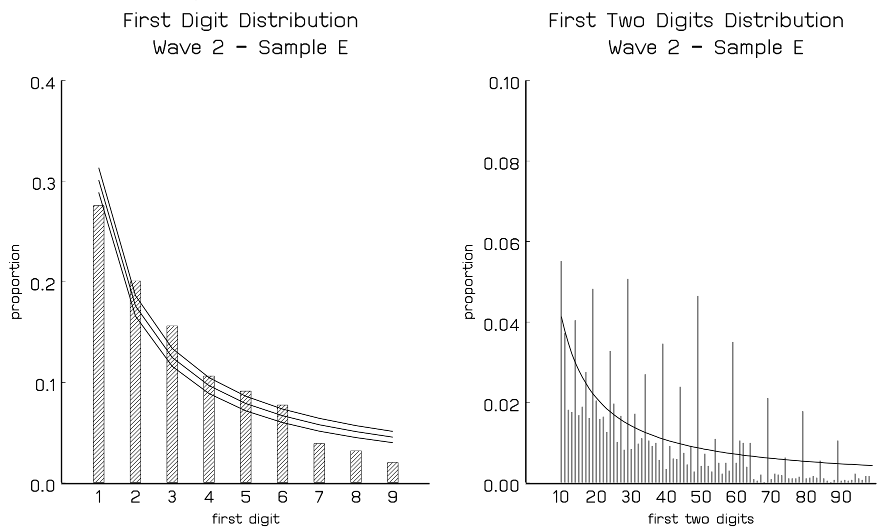


FIGURE 16: *Sample E, wave 2, $\chi^2 = 246$, GFI = 0.998 for first digit distribution, $N = 5,568$, only values with min. 2 digits*

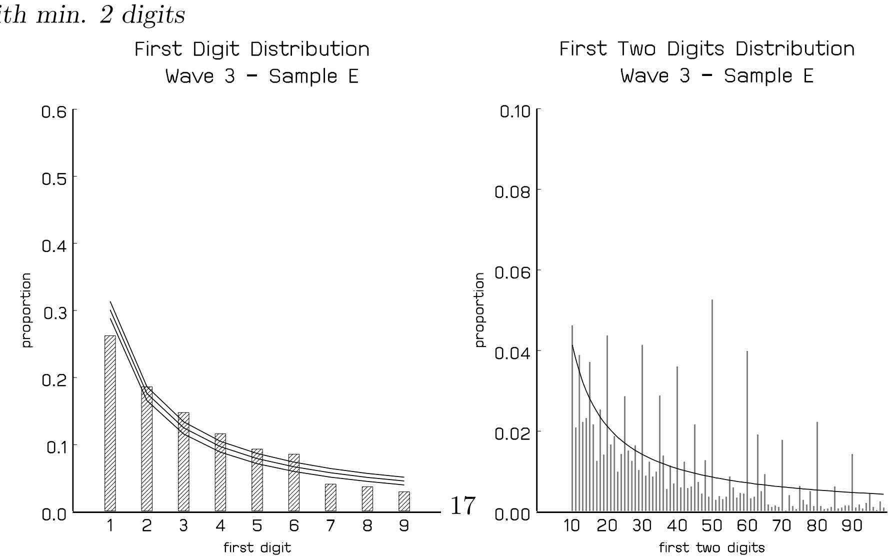


FIGURE 17: *Sample E, wave 3, $\chi^2 = 246$, GFI = 0.998 for first digit distribution, $N = 5,173$, only values with min. 2 digits*

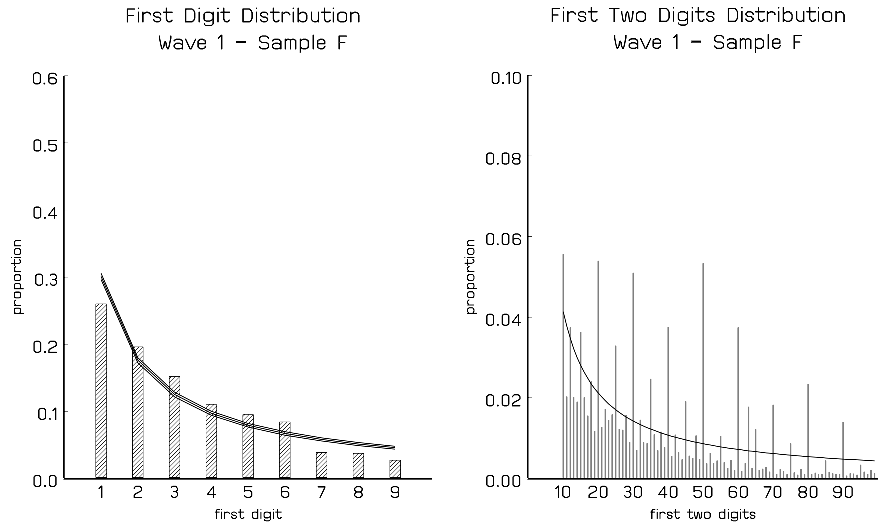


FIGURE 18: *Sample F, wave 1, $\chi^2 = 1,550$, GFI = 0.998 for first digit distribution, $N = 37,656$, only values with min. 2 digits*

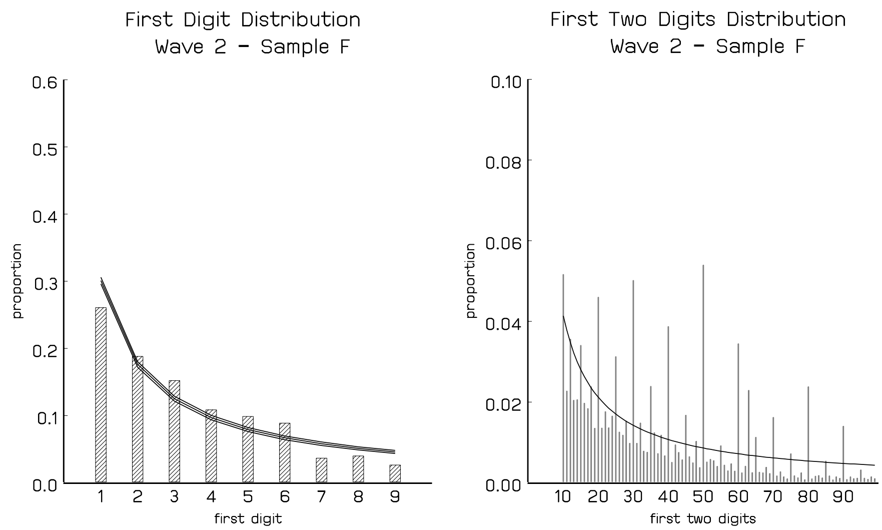


FIGURE 19: *Sample F, wave 2, $\chi^2 = 1,320$, GFI = 0.998 for first digit distribution, $N = 31,910$, only values with min. 2 digits*

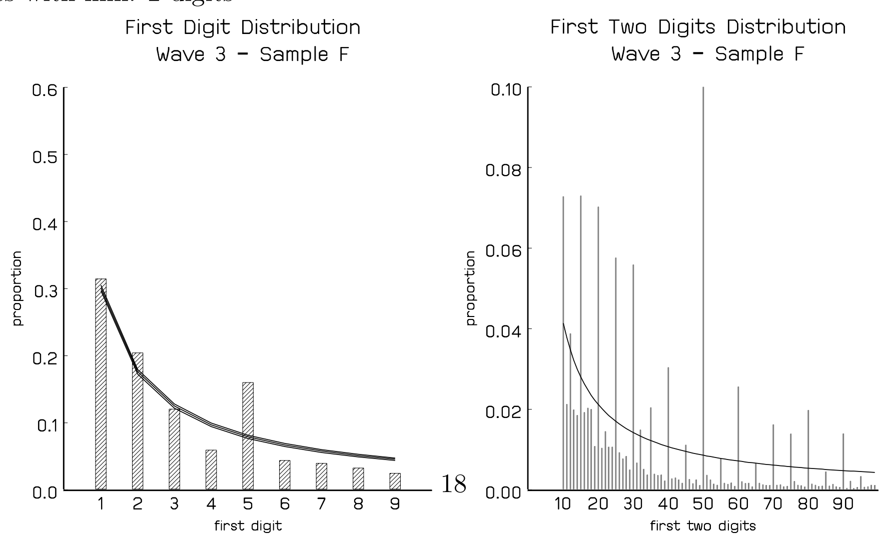


FIGURE 20: *Sample F, wave 3, $\chi^2 = 1,830$, GFI = 0.998 for first digit distribution, $N = 47,140$, only values with min. 2 digits*

5 Identifying interviewer clusters with unusual patterns in relation to Benford's Law

In contrast to cross-sectional surveys, falsification is extremely difficult in complex long-term panel studies like the SOEP because the respondent is mainly interviewed face to face every year, and regular consistency checks between waves show irregularities in the data immediately. Hence, we can assume that fabricated data would generally be a problem only in the first wave and would be detected quite quickly after conducting the second wave. We therefore focus our analysis on the first, second, and third waves of several SOEP subsamples.

For testing the Benford Law procedures, we obtained (true) falsified records from the fieldwork organization that were previously detected using several conventional verification methods and statistical tests of stability and consistence (see Schräpler/Wagner 2005). Fabricated data are rare in the SOEP and have always been found in the first wave of each sample (with the exception of the East German sample C and the small sample D, which are considered as 'clean'). Only one interviewer was able to fabricate data for the first two waves without raising suspicion until wave 3 (sample E). The first wave of samples A and B only contains 0.6 and 1.5% fabricated data, respectively, and the first wave of sample E contains about 2% falsified household interviews. In the second wave approximately 1% of fabricated data was identified in sample E. In the first wave of sample F only 0.1% of the interviews were detected as fabricated. Schräpler and Wagner (2005) have shown that the interviewers who fabricate data usually fabricate a large proportion of their assignment. It therefore gives more statistical power if we analyze whole clusters of interviews per interviewer rather than individual questionnaires. If real survey data follow the logarithmic distribution and fabricated survey data do not, we should be able to identify these clusters of fabricated interviews and test them for significance.

Hence, we now inspect the fit in all interviewer clusters to detect clusters with 'unusual patterns'. We count the first digits to get the digit distribution in each interviewer cluster. If the data from each field representative is viewed as arising from a random sample, we again use the Pearson's chi-square test statistic as a starting point in determining whether an interviewer has collected data following Benford's Law:

$$\chi_i^2 = n_i \sum_{d=1}^{9} \frac{(h_{d_i} - h_{b_d})^2}{h_{b_d}} \tag{5}$$

where n_i is the number of first digits in the interviewer cluster i, h_{d_i} is the observed proportion of digit $d = 1, \ldots, 9$ in interviewer cluster i and h_{b_d} is the proportion of digit d under Benford's distribution.

As already mentioned above, the usage of Pearson's chi-square statistic has the disadvantage that the values depend partly on the number of observations. We will, hence, get higher chi-square values for the same deviations if some interviewer clusters have more digits than others. This makes a comparison of clusters complicated. The other measurement used, GFI, in the section before, was not a good alternative either because the values obtained were, in all cases, quite close to one. It is therefore necessary to develop a better test technique to ensure that all interviewer clusters can be compared. A better solution could be the calculation of probabilities for the chi-square values based on a resampling method like a bootstrap.

5.1 Plausibility values using a resampling method

The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates. Bootstrap samples are generated by resampling with replacement B times from the original data set. For instance, with $n = 6$ we might obtain $\mathbf{x}^* = (\mathbf{x_5}, \mathbf{x_3}, \mathbf{x_5}, \mathbf{x_4}, \mathbf{x_6}, \mathbf{x_1})$. The bootstrap algorithm begins by generating from a large number B of independent bootstrap samples $x^{*1}, x^{*2}, \ldots, x^{*B}$, each of size n.

Then we get bootstrap replicates by calculating the value of the statistic $\hat{\theta}^*(b)$ on each bootstrap sample x^{*b}. If $\hat{\theta}$ is the sample chi-square value to Benford, for instance, then $\hat{\theta}^*(b)$ is the chi-square value to Benford of the bootstrap sample b. More formally, the algorithm for the non-parametric bootstrap is as follows:

1. Sample n observations randomly with replacement from $\mathbf{x_{obs}}$ to obtain a bootstrap data set, denoted $\mathbf{X}^*$.
2. Calculate the bootstrap version of the statistic of interest, $\hat{\theta}^* = \hat{\theta}(\mathbf{X}^*)$
3. Repeat steps 1 and 2 a large number of times, say B, to obtain an estimate of the bootstrap distribution.

In our specific case, the statistic of interest is the chi-square value of an interviewer cluster with size n. We intend to find the probability for the realized or more extreme chi-square value of an interviewer cluster with a certain size of n digits.

A key question is how large B should be. Whereas for standard errors $B = 50$ is often enough to give a good estimate of $se(\hat{\theta})$ much bigger values of B are required for bootstrap confidence intervals (Efron/Tibshirani 1993, p.52). For 90-95 percent confidence intervals Efron and Tibshirani (1993, p.162) suggest that B should be 1,000 or more. As we intend to estimate probabilities, we choose for B at least 2,000.

Probability based on standard normal theory Suppose we obtain our data by random sampling from an unknown distribution F,$F \to x = (x_1, x_2, \ldots, x_n)$. Let $\hat{\theta}^*$ be the estimate of a parameter of interest $\theta = t(F)$, and let $\hat{se}$ be a reasonable estimate of standard error for $\hat{\theta}$, based on bootstrap computations. Under most circumstances, we find that, as the sample size n grows larger, the distribution of $\hat{\theta}$ becomes more and more normal, with mean near θ and variance near $\hat{se}^2$, such that we can assume that asymptotically

$$\frac{\hat{\theta} - \theta}{\hat{se}} \sim N(0, 1) \tag{6}$$

and from there we can calculate an approximation for the observed significance level of an estimator, respectively the probability of obtaining a value of test statistic (here the chi-square value of an interviewer cluster) more extreme than that actually observed $Prob(\theta > \hat{\theta})$

$$P(norm) = Prob(\theta > \hat{\theta}) = 1 - \Phi(\frac{\hat{\theta} - \theta}{\hat{se}}) \tag{7}$$

Of course equation 7 is only an approximation and works well if the bootstrap distribution of $\hat{\theta}^*$ is roughly normal.

Percentile interval method The central limit theorem tells us that as $n \to \infty$, the bootstrap histogram will become normal shaped, but for small samples it may look very abnormal. In this case there is good reason to choose the percentile interval method. This method uses the percentiles of the bootstrap histogram to define confidence limits and significance tests. Again we generate B independent bootstrap data sets $x^{*1}, x^{*1}, \ldots, x^{*B}$ for each interviewer cluster with size n (number of digits in the cluster) and compute (for the chi-square statistic) bootstrap replications $\hat{\theta}^*(b),\quad b = 1, 2, \ldots, B$. Let $\hat{\theta}_B^{*(\alpha)}$ be the $100 \cdot \alpha$th empirical percentile of $\hat{\theta}(b)$ values, that is the $B \cdot \alpha$th value in the ordered list of the B replication of $\hat{\theta}^*$. If $B = 1,000$ and $\alpha = .05$, $\hat{\theta}_B^{*(\alpha)}$ is the 50th ordered value of the replications. Analogue $\hat{\theta}_B^{*(1-\alpha)}$ is the $100 \cdot (1 - \alpha)$th empirical percentile (cf. Efron/Tibshirani (1993, p.170)).

Besides percentile intervals, an approximation of the probability of obtaining a value of test statistic (chi-square values) more extreme than that actually observed $Prob(\theta > \hat{\theta})$ can be obtained directly from the proportion of bootstrap replications higher than the original estimate $\hat{\theta}$

$$P(perc) = Prob(\theta > \hat{\theta}) = 1 - (\frac{\#\hat{\theta}^*(b) < \hat{\theta}}{B}) \tag{8}$$

So, with both methods, we can achieve the normal standard and the percentile interval method probability values for the original chi-square values of the interviewer clusters that are independent of the size of the interviewer clusters. These probabilities reflect the *plausibility* of the fit to Benford, independent of the number of digits in the cluster.

Our hypothesis is that cheating interviewers will have very low probabilities. Hence, it might be useful to construct interviewer rankings by plausibility values.

Interviewer ranking by plausibility We now have to decide which method of probability calculation will be the best for our problem. To find an answer it might be useful to look at the distribution of the bootstrap statistic. As an example, figure 21 shows the distribution of the bootstrap chi-square values calculated on the basis of 42 digits and 1,000 replications for sample E, wave 1. Although we use $B = 1,000$ the shape of the graph is left-skewed and does not really look normal.

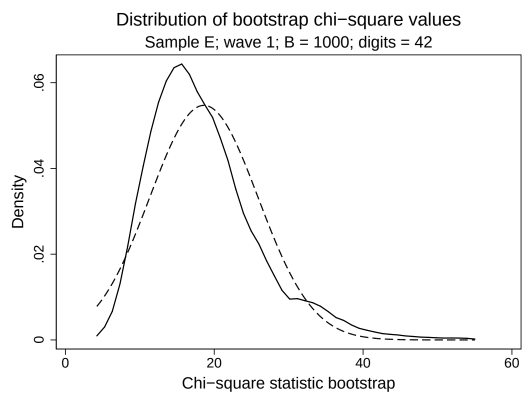


Figure 21: *Bootstrap chi-square values (normal density dotted line)*

The mean of the bootstrap chi-square distribution is 18.612 and the standard deviation 7.29. The realized chi-square value for the interviewer is 11.186 in our example and lower than the mean. The probability of obtaining a chi-square value more extreme than $Prob(\theta > \hat{\theta} = 11.186)$ based on normal theory is, according to equation 7

$$P(norm) = Prob(\theta > 11.186) = 1 - \Phi(\frac{11.186 - 18.612}{7.29}) = 0.8458$$

and analogous to the probability based on the percentile interval method using equation 8

$$P(perc) = Prob(\theta > 11.186) = 1 - (\frac{125}{1000}) = 0.875$$

where 125 realized values of the 1,000 bootstrap replications are lower than the value 11.186. We can see that the probability value of the percentile interval method is higher than the value that is obtained by normal theory because the first takes account of the fact that the median is lower than the average mean of the bootstrap chi-square distribution. It therefore seems to be reasonable to use the percentile interval method to calculate probabilities for each interviewer cluster in our application.[13]

5.2 Fit in interviewer clusters of sample A/B

The scatterplots in figures 22-27 show the fit to Benford for the first digit and first two digit distribution in each interviewer cluster in samples A/B.[14] The chi-square values of the clusters with detected fabricated interviews are marked with black circles. We can see that one of the four fabricated clusters has the worst fit to Benford and appears as an outlier in the case of the first digit distribution. In the first two digit distribution, three of the marked clusters have very high fit values.

Figures 28 and 29 on page 24 show the density distribution[15] of the probability $P(perc)$ in samples A/B, wave 1 for the first digit and first two digit distribution (normal density dotted line). If all interviewers are free from suspicion, $P(perc)$ would only have values above 0.5 and the density function would ideally have a peak near $P(perc) = 1.0$. In our case, the highest density occurs at $P(perc) = 0.94$. Furthermore we can also see in the low probability region at $P(perc) = 0.1$ a local maximum. This means that there are a number of clusters with very low plausible fit values. One reason might be that these interviewers work in quite homogeneous sample points and/or that some of these interviewers fabricate their assignment and fail Benford's Law.

Table 2 on page 24 shows the interviewer-ranking by the probability $P(perc)$ of each cluster for wave 1 to 3, sample A/B. We can see that the fabricated cluster of interviewer already identified, No. xx827x with 122 digits, has the lowest probability $P(perc) = 0.002$ of all interviewers in wave 1. Overall we find six additional interviewers who have probabilities below the 5% level. Of course, this is not a sure indication that these clusters are fabricated but low plausibilities for the realized chi-square values could be a result of cheating and the fieldwork organization can use this information to recontact households in suspicious interviewer clusters.

Unfortunately the two other fakes evident in wave 1 could not be identified with the first digit distribution. The cheating interviewer No. xx800x has rank 61 (P(perc) = 0.265) and interviewer No. xx937x even has a really high plausibility of 0.958 and rank 420 (not shown in the table). Nevertheless, if we use the first two digit Benford distribution, we will find three of four cheating interviewers in the top 12 of the ranking list, shown in table 20 on page 55 in the appendix. This indicates that, in some cases, the first two digit distribution is more successful.

[13] The probability calculations are done with the GAUSS Programming Language (Aptech Systems, Inc.)
[14] All scatterplots in this paper are done with the software program TDA (Rohwer/Pötter 2005).
[15] We again use a kernel density estimation with an Epanechnikov kernel.

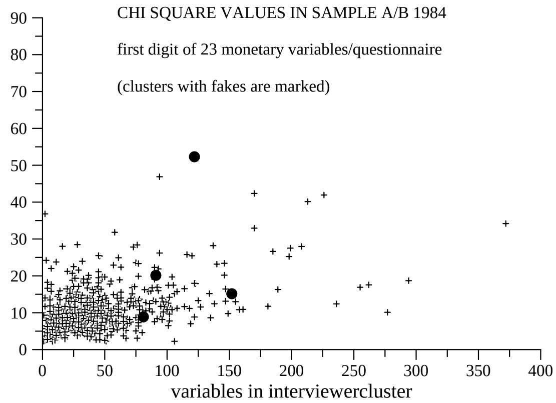


FIGURE 22: *First digit distribution: Chi-Square values for interviewer cluster in wave 1, sample A/B*

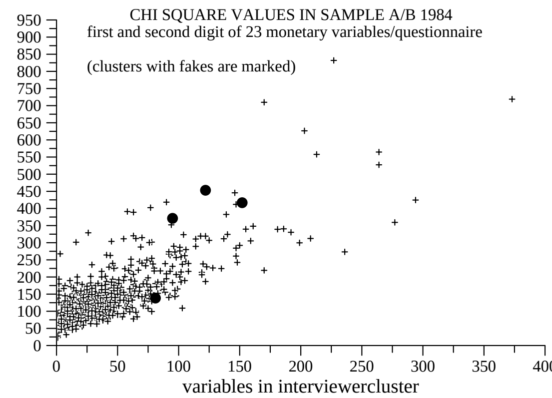


FIGURE 23: *First two digit distribution: Chi-Square values for interviewer cluster in wave 1, sample A/B*

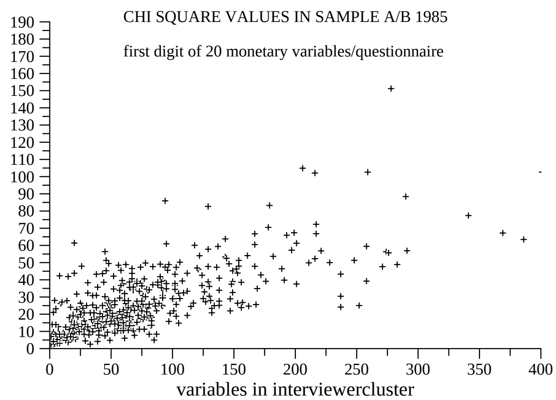


FIGURE 24: *First digit distribution: Chi-Square values for interviewer cluster in wave 2, sample A/B*

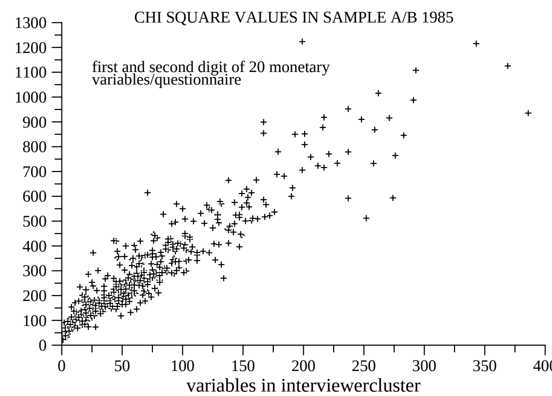


FIGURE 25: *First two digit distribution: Chi-Square values for interviewer cluster in wave 2, sample A/B*

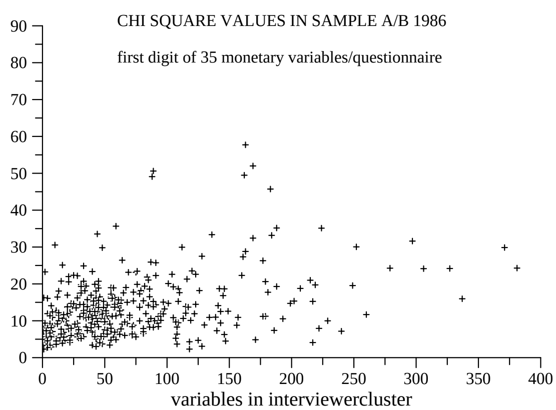


FIGURE 26: *First digit distribution: Chi-Square values for interviewer cluster in wave 3, sample A/B*

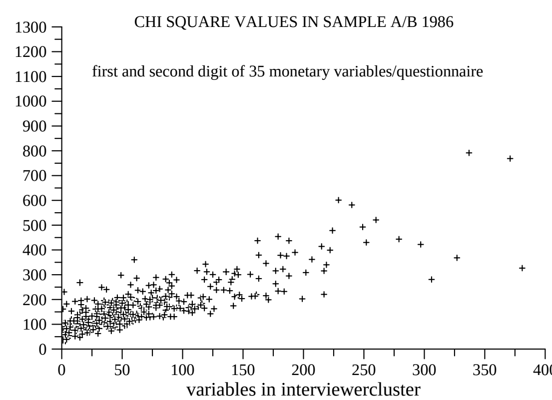


FIGURE 27: *First two digit distribution: Chi-Square values for interviewer cluster in wave 3, sample A/B*

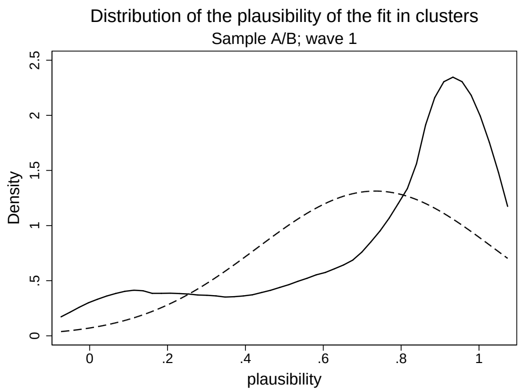


FIGURE 28: *Distribution of the probability $P(perc)$, sample A/B, wave 1, first digits*

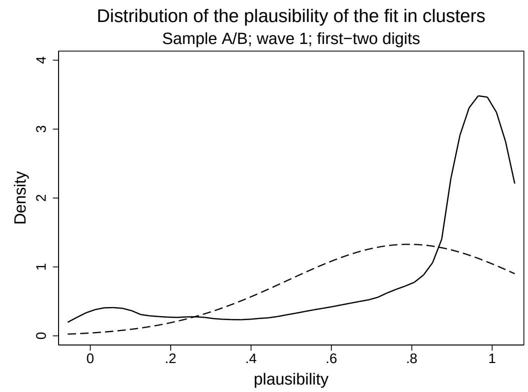


FIGURE 29: *Distribution of the probability $P(perc)$, sample A/B, wave 1, first two-digits*

TABLE 2: *Interviewer-ranking by the probability of the results of each interviewer cluster in wave 1-3, sample A/B (faking interviewer bold), $B = 2,000$*

		wave 1					wave 2					wave 3		
Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)
1	**xx827x**	**122**	**52.30**	**0.0020**	1	xx520x	20	61.39	0.0000	1	xx202x	89	50.61	0.0010
2	xx147x	94	46.88	0.0040	2	xx145x	45	56.37	0.0000	2	xx082x	163	57.72	0.0010
3	xx785x	28	28.48	0.0060	3	xx415x	94	85.91	0.0000	3	xx167x	88	49.10	0.0020
4	xx650x	32	23.95	0.0180	4	xx871x	26	47.92	0.0000	4	xx035x	10	30.56	0.0050
5	xx887x	29	21.56	0.0410	5	xx730x	46	51.29	0.0000	5	xx766x	162	49.47	0.0000
6	xx320x	16	28.01	0.0450	6	xx287x	38	43.28	0.0000	6	xx150x	59	35.65	0.0080
7	xx800x	45	25.50	0.0470	7	xx305x	43	43.56	0.0000	7	xx650x	169	51.99	0.0070
8	xx363x	46	25.37	0.0510	8	xx404x	48	49.47	0.0000	8	xx501x	44	33.54	0.0140
9	xx609x	25	22.51	0.0630	9	xx466x	20	43.82	0.0000	9	xx951x	33	24.89	0.0160
10	xx687x	27	19.34	0.0680	10	xx187x	46	45.31	0.0000	10	xx801x	28	22.19	0.0300
11	xx342x	94	26.19	0.0800	11	xx785x	56	48.50	0.0000	11	xx494x	33	20.72	0.0470
12	xx583x	20	21.22	0.0890	12	xx647x	95	60.93	0.0000	12	xx046x	31	19.80	0.0500
13	xx156x	33	19.18	0.0930	13	xx574x	58	45.45	0.0000	13	xx895x	183	45.73	0.0590
14	xx756x	58	31.81	0.0970	14	xx156x	31	38.30	0.0010	14	xx694x	25	22.34	0.0570
15	xx401x	26	19.35	0.1000	15	xx544x	8	42.32	0.0010	15	xx211x	31	19.30	0.0580
16	xx353x	4	18.24	0.1000	16	xx584x	44	38.60	0.0010	16	xx785x	48	29.81	0.0810
17	xx752x	24	20.69	0.1020	17	xx263x	62	48.88	0.0010	17	xx263x	40	23.35	0.0880
18	xx208x	33	18.62	0.1040	18	xx207x	67	46.57	0.0010	18	xx988x	31	17.54	0.0870
19	xx654x	226	41.93	0.1040	19	xx047x	67	43.73	0.0010	19	xx445x	64	26.43	0.0980
20	xx263x	36	19.09	0.1080	20	xx851x	31	32.34	0.0020	20	xx743x	112	29.95	0.1360
21	xx846x	33	18.43	0.1090	21	xx772x	39	35.72	0.0020	21	xx277x	30	17.47	0.1110
22	xx187x	33	18.09	0.1190	22	xx237x	129	82.65	0.0020	22	xx570x	34	18.14	0.1260
23	xx084x	11	23.76	0.1200	23	xx570x	52	42.15	0.0020	23	xx588x	21	22.00	0.1290
24	xx508x	37	20.14	0.1220	24	xx518x	22	31.72	0.0040	24	xx985x	16	25.11	0.1280
25	xx676x	170	42.35	0.1260	25	xx948x	78	49.77	0.0040	25	xx163x	136	33.36	0.1520
26	xx136x	45	21.13	0.1340	26	xx543x	15	41.89	0.0060	26	xx382x	76	23.45	0.1630
27	xx106x	7	22.00	0.1380	27	xx807x	38	30.94	0.0060	27	xx624x	87	25.92	0.1450
28	xx200x	37	19.50	0.1430	28	xx018x	4	28.02	0.0060	28	xx156x	35	19.41	0.1590
29	xx665x	29	17.15	0.1440	29	xx810x	90	49.10	0.0070	29	xx901x	69	23.15	0.1650
30	xx305x	24	18.81	0.1540	30	xx674x	35	30.78	0.0080	30	xx268x	28	16.20	0.1650
31	xx866x	61	24.91	0.1570	31	xx709x	74	47.34	0.0080	31	xx340x	21	20.56	0.1750
32	xx544x	76	28.40	0.1660	32	xx766x	67	40.64	0.0090	32	xx514x	75	23.14	0.1670
33	xx519x	93	21.89	0.1740	33	xx343x	84	47.67	0.0130	33	xx450x	169	32.42	0.1950
34	xx216x	45	19.50	0.1810	34	xx167x	97	48.90	0.0130	34	xx237x	128	27.51	0.2160
35	xx766x	37	18.11	0.1820	35	xx446x	65	38.70	0.0140	35	xx827x	91	25.73	0.2370
36	xx020x	7	17.68	0.1850	36	xx150x	59	39.60	0.0150	36	xx851x	44	19.59	0.2230
37	xx167x	90	22.30	0.1910	37	xx593x	30	25.00	0.0170	37	xx174x	163	28.80	0.2440
38	xx118x	50	19.68	0.1920	38	xx282x	45	30.36	0.0190	38	xx674x	33	14.43	0.2280
39	xx778x	137	28.20	0.1960	39	xx568x	58	36.73	0.0200	39	xx910x	57	18.92	0.2460
40	xx884x	7	15.79	0.1990	40	xx716x	45	30.16	0.0210	40	xx419x	42	19.88	0.2620
		⋮					⋮					⋮		
61	**xx800x**	**91**	**20.14**	**0.2650**	61	xx582x	103	47.25	0.0410	61	xx373x	120	23.53	0.3930
		⋮					⋮					⋮		
636	xx745x	676	13.37	1.0000	463	xx895x	252	25.02	1.0000	407	xx377x	3	7.31	1.0000

Source SOEP, individual questionnaire, only monetary variables, 1984 - 1986 (own calculation)

5.3 Fit in interviewer clusters of sample C

We have shown in section 4.1.3 on page 15 that Benford's Law doesn't hold in wave 1 and 2 in the East German sample C. We have found a strong disproportion of the lower digits, probably caused by homogeneous cluster with quite low monetary values. The homogeneity in the data is attributed to the living conditions in East Germany in the year 1989.

If the overall fit in the sample is worst, we can reasonably assume that the fit for most clusters will be worst too. Leading from this, we find in figure 30 rather high chi-square values for clusters in wave 1 of sample C (max. chi-sq. = 112.8; digits = 99). In spite of all this, the fieldwork organization could not identify cheating interviewers in this sample.

Figures 36 and 37 on page 27 show the density distribution of the probability $P(perc)$ in sample C, wave 1 (normal density dotted line) for the first digit and first two digit distribution. The shape of the first digit density function is totally different from figure 28 on page 24. We find the highest density near 0.1 and a local maximum at 0.65. A naive interpretation would be that almost all interviewers are suspect. However, this density shape is, of course, caused by the homogeneity of the interviewer clusters in sample C in the first years after German unification. The distribution of $P(perc)$ for the first digit fit statistic shows that the success of Benford's approach is highly dependent on the requirement that Benford's Law holds for the whole sample.[16] The density distribution for the first two digit fit statistic in figure 37 seems to be more suitable. The shape shows a local maximum near 0.1 and a maximum near 0.98. Table 3 on page 27 shows the interviewer-ranking based on the plausibility of the fit for the first digit distribution for sample C.[17] We can see that in wave 1 approx. 80 interviewers have a value of $P(perc) < 0.05$.

5.4 Fit in interviewer clusters of sample E

The overall fit to Benford in sample E is shown in the figures 15-17 on page 17. Only small variances from the predicted distribution can be observed. We can, therefore, reasonably assume that we can use the logarithmic distribution to detect fabrications. The figures 38-43 on page 28 show the scatterplots of the chi-square values for each interviewer cluster. The marked falsified clusters are obviously outliers in the first digit distribution.

Figure 44 on page 29 shows the density distribution of the probability $P(perc)$ in sample E, wave 1 (normal density dotted line). Because sample E contains only 1,957 respondents (including fakes) we are able to use a good deal more bootstrap replications with $B = 10,000$ than in sample A/B ($B = 2,000$) without encountering computational problems. The shape of the density distribution is very similar to the distribution of sample A/B (figure 28 on page 24). We find the highest density near value 0.95 and a local maximum at value 0.15. Most clusters therefore have very plausible fits to the logarithmic distribution. The shape seems suitable for detecting fraudulent interviewers.

Table 4 on page 29 shows the interviewer ranking for sample E. Fraudulent interviewers are framed and marked in bold. We find three of five cheating interviewers within the top 7. Furthermore, the interviewer who faked two waves is at the top of the list in wave 2. The two undetected cheating interviewers have only one (fabricated) personal interview each. We can assume that this cluster size is too small for our detection procedure. Their positions in the ranking list are therefore 118 and 69, respectively. However, overall our empirical results show that Benford's approach is remarkably successful in the case of sample E.

[16] In section 5.7 on page 33 we introduce an alternative, more general procedure, that doesn't assume that Benford's Law holds exactly for a particular data set. The only assumption is that the vast majority of interviewers are honest. This alternative should perform better in the case of sample C.

[17] The ranking for the first two digit distribution is shown in the appendix in table 21 on page 55.

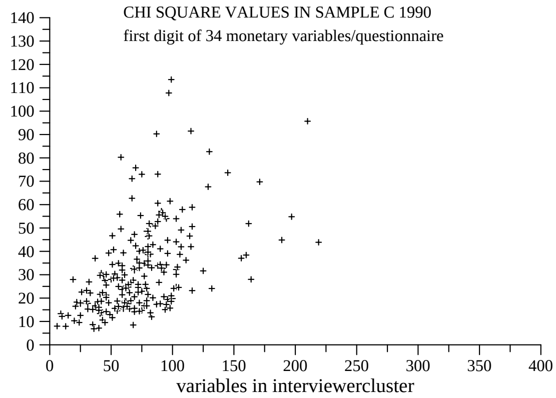


FIGURE 30: *First digit distribution: Chi-Square values for interviewer cluster in sample C wave 1*

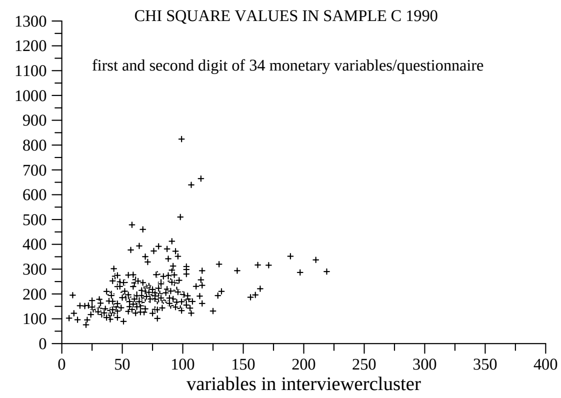


FIGURE 31: *First-two digit distribution: Chi-Square values for interviewer cluster in sample C wave 1*

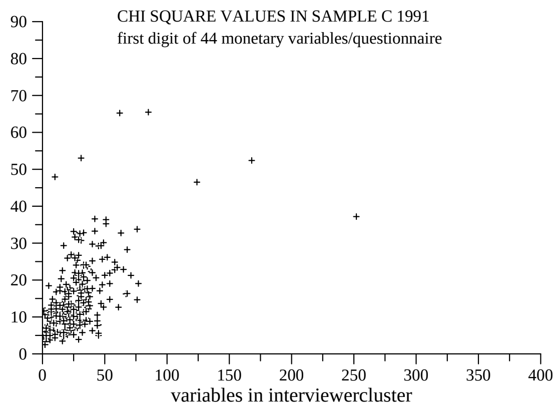


FIGURE 32: *First digit distribution: Chi-Square values for interviewer cluster in sample C wave 2*

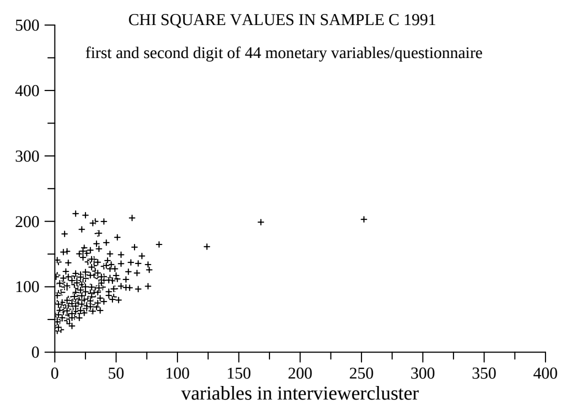


FIGURE 33: *First-two digit distribution: Chi-Square values for interviewer cluster in sample C wave 2*

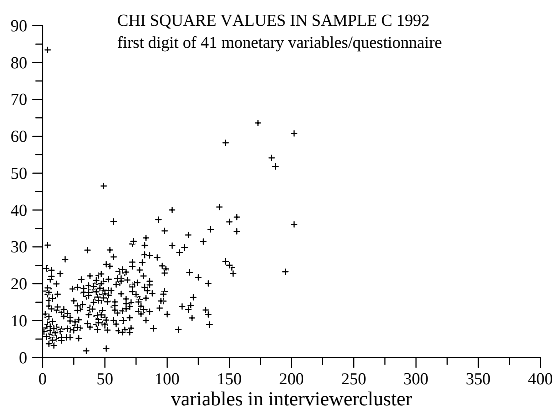


FIGURE 34: *First digit distribution: Chi-Square values for interviewer cluster in sample C wave 3*

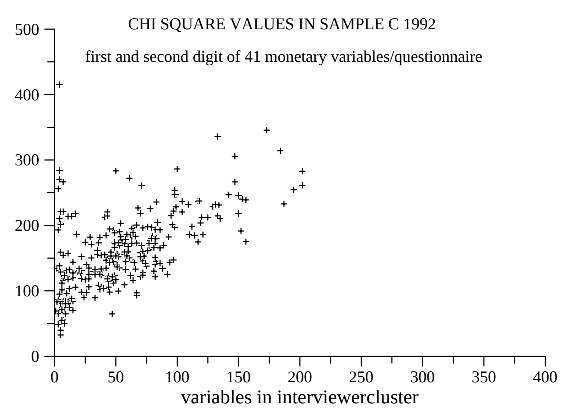


FIGURE 35: *First-two digit distribution: Chi-Square values for interviewer cluster in sample C wave 3*

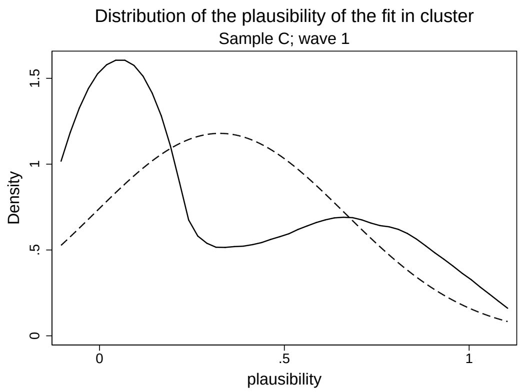


FIGURE 36: *Distribution of the probability $P(perc)$, sample C, wave 1, first digits*

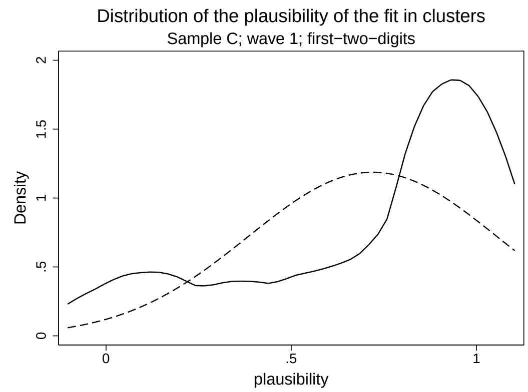


FIGURE 37: *Distribution of the probability $P(perc)$, sample C, wave 1, first two-digits*

TABLE 3: *Interviewer ranking by the plausibility of the Interviewer clusters in wave 1-3, sample C, $(B = 10,000)$*

		wave 1					wave 2					wave 3		
Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)
1	xx246x	58	79.81	0	1	xx246x	10	47.71	0	1	xx177x	147	58.24	0
2	xx530x	88	72.97	0	2	xx327x	62	65.23	0	2	xx264x	4	83.42	0
3	xx611x	99	112.79	0	3	xx303x	31	52.89	0	3	xx929x	313	99.12	0
4	xx645x	67	70.60	0	4	xx960x	5	18.46	0	4	xx664x	173	63.67	0.0001
5	xx840x	87	89.84	0	5	xx323x	85	65.46	0.0005	5	xx452x	93	37.38	0.0002
6	xx056x	67	62.62	0	6	xx754x	17	29.13	0.0010	6	xx323x	202	60.71	0.0005
7	xx338x	97	107.67	0	7	xx326x	25	33.17	0.0010	7	xx377x	104	40.05	0.0005
8	xx750x	75	72.99	0	8	xx213x	33	32.73	0.0011	8	xx248x	142	40.87	0.0014
9	xx553x	70	75.77	0	9	xx783x	20	25.95	0.0016	9	xx410x	187	51.86	0.0015
10	xx670x	57	55.83	0.0001	10	xx331x	30	32.60	0.0020	10	xx611x	184	54.17	0.0016
11	xx053x	115	91.40	0.0002	11	xx803x	42	36.58	0.0027	11	xx254x	49	46.53	0.0023
12	xx111x	51	46.56	0.0002	12	xx172x	26	32.86	0.0029	12	xx114x	83	32.46	0.0023
13	xx884x	74	55.31	0.0003	13	xx010x	26	31.66	0.0036	13	xx771x	82	30.41	0.0025
14	xx800x	130	82.58	0.0004	14	xx354x	29	30.92	0.0045	14	xx932x	98	34.36	0.0028
15	xx706x	88	60.21	0.0007	15	xx721x	51	36.38	0.0047	15	xx800x	104	30.15	0.0037
16	xx811x	92	56.25	0.0009	16	xx451x	31	30.72	0.005	16	xx606x	117	33.27	0.0039
17	xx393x	80	48.56	0.0011	17	xx550x	16	22.57	0.0053	17	xx196x	135	34.77	0.0051
18	xx264x	91	56.86	0.0012	18	xx800x	51	35.24	0.0066	18	xx371x	110	28.49	0.0054
19	xx050x	58	49.60	0.0012	19	xx460x	42	33.21	0.0073	19	xx427x	82	27.97	0.0058
20	xx211x	91	56.90	0.0012	20	xx766x	24	28.11	0.0080	20	xx102x	82	27.78	0.0062
21	xx303x	98	61.05	0.0012	21	xx452x	25	26.52	0.0082	21	xx706x	57	36.89	0.0063
22	xx261x	89	55.58	0.0014	22	xx932x	40	29.75	0.0090	22	xx238x	18	26.61	0.0067
23	xx118x	79	48.60	0.0016	23	xx498x	33	24.11	0.0157	23	xx796x	92	27.15	0.0068
24	xx561x	81	51.72	0.0016	24	xx125x	26	25.98	0.0159	24	xx164x	150	36.83	0.0074
25	xx076x	80	46.26	0.0018	25	xx421x	33	23.92	0.0171	25	xx326x	73	31.55	0.0080
26	xx170x	103	53.70	0.0021	26	xx822x	29	26.71	0.0178	26	xx921x	114	29.84	0.0087
27	xx164x	108	57.83	0.0023	27	xx121x	15	20.34	0.0213	27	xx498x	80	25.76	0.0104
28	xx248x	95	54.03	0.0024	28	xx962x	15	20.33	0.0219	28	xx617x	72	30.81	0.0106
29	xx220x	52	40.73	0.0027	29	xx750x	19	18.82	0.0234	29	xx961x	156	38.12	0.0107
30	xx571x	88	52.56	0.0029	30	xx480x	28	25.37	0.0247	30	xx770x	86	27.72	0.0112
31	xx622x	94	54.72	0.0030	31	xx584x	49	30.15	0.0279	31	xx131x	129	31.10	0.0115
32	xx737x	81	46.48	0.0037	32	xx828x	35	24.13	0.0279	32	xx326x	96	24.87	0.0158
33	xx261x	129	67.64	0.0038	33	xx751x	45	29.22	0.0297	33	xx806x	78	23.72	0.0245
34	xx452x	69	47.25	0.0041	34	xx024x	52	26.16	0.0324	34	xx931x	99	24.06	0.0270
35	xx121x	82	51.59	0.0042	35	xx664x	63	32.74	0.0326	35	xx665x	156	34.16	0.0285
36	xx070x	116	58.42	0.0052	36	xx126x	40	25.17	0.0332	36	xx256x	99	23.80	0.0290
37	xx766x	80	42.01	0.0053	37	xx780x	36	23.72	0.0343	37	xx125x	54	29.17	0.0300
38	xx109x	86	50.88	0.0056	38	xx528x	27	23.97	0.04	38	xx070x	81	22.14	0.0377
39	xx151x	48	39.22	0.0062	39	xx335x	19	16.79	0.0407	39	xx338x	72	25.91	0.0384
40	xx617x	76	40.37	0.0071	40	xx611x	33	20.86	0.0409	40	xx941x	98	22.92	0.0396
		⋮					⋮					⋮		
214	xx091x	20	10.31	1.0000	264	xx508x	3	4.99	1.0000	278	xx983x	8	4.71	1.0000

Source SOEP, individual questionnaire, only monetary variables, 1989 - 1991 (own calculation)

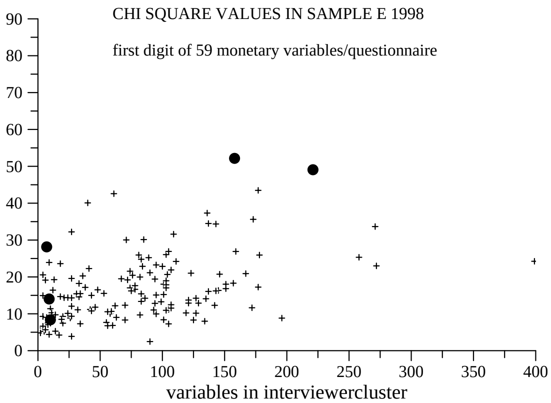


FIGURE 38: *First digit distribution: Chi-Square values for interviewer cluster in wave 1, sample E*

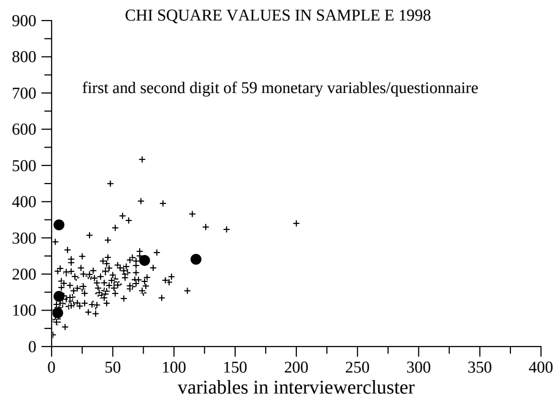


FIGURE 39: *First two digit distribution: Chi-Square values for interviewer cluster in wave 1, sample E*

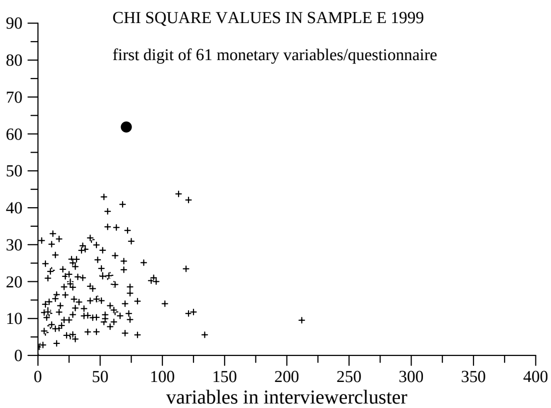


FIGURE 40: *First digit distribution: Chi-Square values for interviewer cluster in wave 2, sample E*

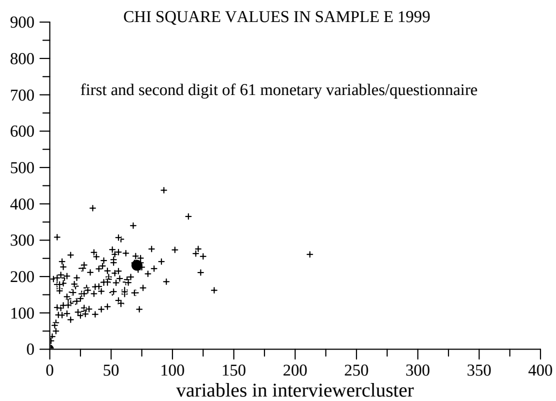


FIGURE 41: *First two digit distribution: Chi-Square values for interviewer cluster in wave 2, sample E*

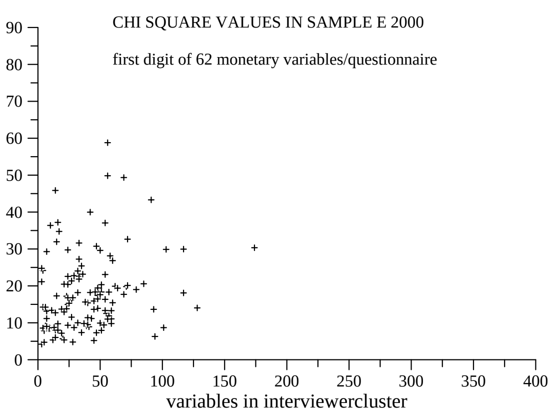


FIGURE 42: *First digit distribution: Chi-Square values for interviewer cluster in wave 3, sample E*

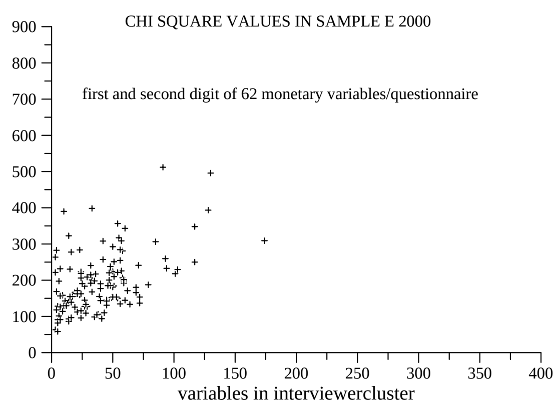


FIGURE 43: *First two digit distribution: Chi-Square values for interviewer cluster in wave 3, sample E*

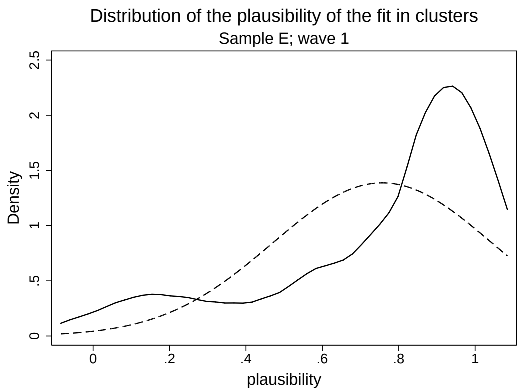


FIGURE 44: *Distribution of the probability $P(perc)$, sample E, wave 1, first digits*

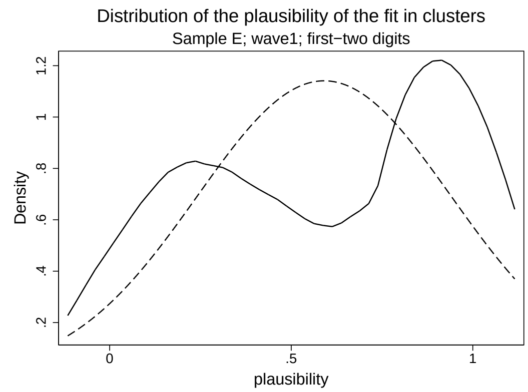


FIGURE 45: *Distribution of the probability $P(perc)$, sample E, wave 1, first two-digits*

TABLE 4: *Interviewer ranking by plausibility of Interviewer clusters in wave 1-3, sample E, ($B = 10,000$)*

Rank	Intnr	wave 1 digits	chi-sq.	P(perc)	Rank	Intnr	wave 2 digits	chi-sq.	P(perc)	Rank	Intnr	wave 3 digits	chi-sq.	P(perc)
1	**xx683x**	**221**	**49.07**	**0.0016**	1	**xx928x**	**71**	**61.88**	**0**	1	xx679x	130	98.24	0
2	xx232x	61	42.58	0.0126	2	xx671x	68	40.92	0.0046	2	xx589x	56	58.79	0
3	xx066x	40	40.08	0.0169	3	xx868x	53	42.94	0.0071	3	xx022x	56	49.83	0.0005
4	**xx928x**	**158**	**52.16**	**0.0221**	4	xx690x	17	31.55	0.0087	4	xx724x	69	49.35	0.0007
5	xx679x	177	43.48	0.0427	5	xx242x	121	42.11	0.0097	5	xx202x	54	37.05	0.0039
6	xx690x	27	32.22	0.1028	6	xx679x	113	43.75	0.0114	6	xx232x	33	31.62	0.0097
7	**xx928x**	**7**	**28.15**	**0.1126**	7	xx801x	72	33.86	0.0178	7	xx226x	42	39.98	0.0128
8	xx469x	173	35.62	0.1591	8	xx282x	12	32.99	0.0179	8	xx446x	33	27.26	0.0236
9	xx674x	85	30.14	0.1662	9	xx792x	36	29.70	0.0210	9	xx278x	35	25.41	0.0354
10	xx905x	18	23.60	0.1746	10	xx827x	47	29.92	0.0255	10	xx868x	50	29.61	0.0477
11	xx708x	136	37.32	0.1843	11	xx761x	75	30.93	0.0288	11	xx389x	32	24.04	0.0524
12	xx370x	271	33.66	0.2140	12	xx923x	38	28.75	0.0289	12	xx681x	33	22.62	0.0593
13	xx761x	143	34.36	0.2237	13	xx933x	20	23.33	0.0304	13	xx433x	36	23.17	0.0598
14	xx318x	71	30.04	0.2402	14	xx963x	11	30.13	0.0308	14	xx858x	33	21.84	0.0680
15	xx690x	137	34.49	0.2466	15	xx527x	28	25.06	0.0311	15	xx134x	47	30.75	0.0741
16	xx933x	41	22.28	0.2759	16	xx232x	35	28.46	0.0326	16	xx761x	72	32.66	0.0764
17	xx037x	89	25.23	0.2870	17	xx759x	27	26.07	0.0341	17	xx107x	54	23.09	0.0803
18	xx589x	109	31.60	0.2934	18	xx589x	63	34.65	0.0348	18	xx923x	24	29.74	0.0841
19	xx693x	9	23.92	0.3316	19	xx674x	22	21.42	0.0400	19	xx651x	58	28.14	0.0919
20	xx660x	258	25.33	0.3350	20	xx234x	56	39.03	0.0412	20	xx293x	14	45.86	0.0942
21	xx242x	13	19.27	0.3639	21	xx833x	43	31.37	0.0449	21	xx489x	29	22.78	0.0953
22	xx544x	83	24.78	0.4486	22	xx251x	30	24.05	0.0454	22	xx568x	29	22.59	0.0992
23	xx076x	178	25.93	0.4665	23	xx350x	31	26.04	0.0471	23	xx690x	91	43.31	0.1032
24	xx553x	105	26.91	0.4704	24	xx127x	14	27.21	0.0534	24	xx686x	16	37.20	0.1205
25	xx268x	81	25.95	0.4830	25	xx207x	42	31.83	0.0557	25	xx568x	17	34.75	0.1247
26	xx881x	90	21.15	0.5027	26	xx638x	25	21.98	0.0648	26	xx469x	60	26.85	0.1258
27	xx739x	84	22.87	0.5099	27	xx878x	21	18.56	0.0703	27	xx691x	32	18.19	0.1615
28	xx990x	103	26.05	0.5166	28	xx668x	56	34.83	0.0759	28	xx530x	27	21.34	0.1745
29	xx811x	159	26.91	0.5281	29	xx202x	69	25.54	0.0840	29	xx150x	24	22.58	0.2163
30	xx278x	111	24.20	0.5459	30	xx895x	52	28.49	0.0858	30	xx066x	51	20.33	0.2190
31	xx317x	95	23.25	0.5573	31	xx686x	11	23.05	0.0951	31	xx690x	15	31.95	0.2395
32	xx020x	67	19.50	0.5907	32	xx649x	48	25.89	0.0984	32	xx980x	69	17.71	0.2764
33	xx607x	27	19.59	0.5963	33	xx450x	26	19.96	0.1085	33	xx450x	24	20.38	0.2801
34	xx069x	43	14.99	0.5967	34	xx336x	26	19.38	0.1237	34	xx370x	28	16.80	0.2841
35	xx488x	33	18.23	0.5976	35	xx550x	10	22.76	0.1239	35	xx843x	57	18.32	0.2866
36	xx386x	272	23.00	0.6019	36	xx022x	62	27.03	0.1266	36	xx895x	51	18.34	0.2990
37	xx690x	12	16.42	0.6311	37	xx469x	69	23.20	0.1283	37	xx160x	54	16.35	0.3006
38	xx827x	74	21.55	0.6376	38	xx389x	22	16.38	0.1310	38	xx282x	4	24.17	0.3165
39	xx256x	18	14.67	0.6456	39	xx054x	51	23.56	0.1344	39	xx386x	71	19.81	0.3187
40	xx348x	36	20.27	0.6736	40	xx414x	36	21.02	0.1348	40	xx984x	64	19.39	0.3291
		⋮					⋮					⋮		
150	xx239x	6	5.68	1.0000	125	xx607x	4	2.83	1.0000	129	xx972x	12	5.35	1.0000

Source SOEP, sample E, individual questionnaire, only monetary variables, 1998 - 2001 (own calculation)

5.5 Fit in interviewer clusters of sample F

The figures 18 and 19 on page 18 suggest that Benford's Law holds for the first waves of sample F. The empirical distributions show quite a close fit (except wave 3 in figure 20, where the digit 5 has a higher proportion than predicted). The falsified cluster that has already been detected by the fieldwork organization is marked in black in the scatterplots in figures 46 and 47 on page 31. The number of different continuous and monetary variables in each data set is, at approximately 60, quite high. Again, the marked faked cluster seems to be an outlier in the first digit distribution.

Because sample F contains 10,481 respondents (with falsifications) and more than triple the number of interviewer clusters (536) as in sample E, we have to reduce the bootstrap replications to $B = 2,000$ to avoid computation problems. The density distribution of the probability $P(perc)$ in wave 1 is shown in figure 52 on page 32. The shape of this distribution is quite similar to the shape for samples A/B and E. The highest density is, again, close to the value 0.95 and we find a local maximum in the range of 0.15 to 0.4.

Table 5 on page 32 shows the interviewer ranking for the plausibility of the first digit fit statistic values in sample F. The cheating interviewer, who had already been detected, is framed and marked in bold. He is listed within the top ten in the wave 1 list.[18] This indicates that our detection procedure is also successful for sample F.

5.6 Predictive power of Benford - a new falsifier is detected

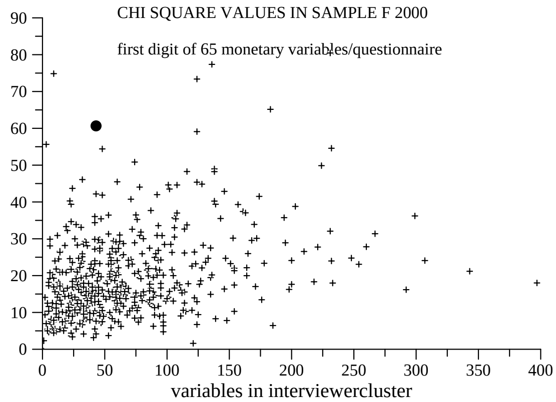
The aim of the study is not only to show that the Benford distribution allows us to identify falsifications that have already been detected. We also intend to detect fabrications that have not yet been found with the conventional quality control methods. We, therefore, now attempt to identify additional fabrications in the survey.

To test the predictive power of Benford's Law, we have consulted the fieldwork organization, Infratest, to check our interviewer ranking lists and match them with their own information. Because the data collections in samples A/B, C, and E were made more than six years ago, we concentrate our inquiry on the newest, subsample F. The first wave of sample F started in the year 2000.

An investigation by Infratest produced an astonishing result: in our list, Interviewer no. xx713x, who had been fired because of unreliability, is ranked above no. xx085x, who had been fired after wave 1 because of falsifying interviews. The results of our Benford analysis suggested that interviewer no. xx713x fabricated his interviews as well. A close inspection conducted thereafter by Infratest showed that only two of his ten declared respondents really exist and are reachable. Infratest and DIW Berlin have now labeled this interviewer to be a falsifier, as detected by our method. The data of this interviewer are deleted from the SOEP.

The success of our Benford analysis strongly suggests that this approach is also useful for other survey data.

[18] The ranking for the second digit distribution is shown in the appendix in table 22 on page 56. The falsified interviewer cluster is only ranked in position 58 here.



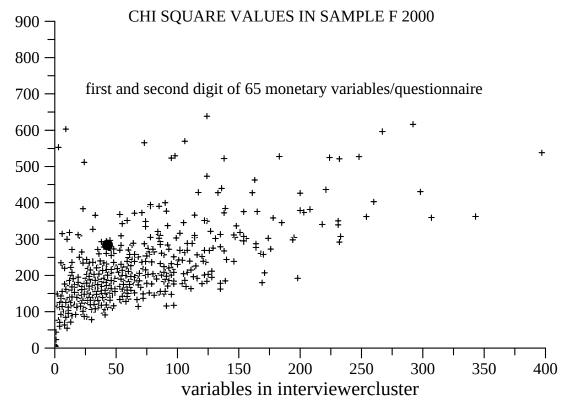
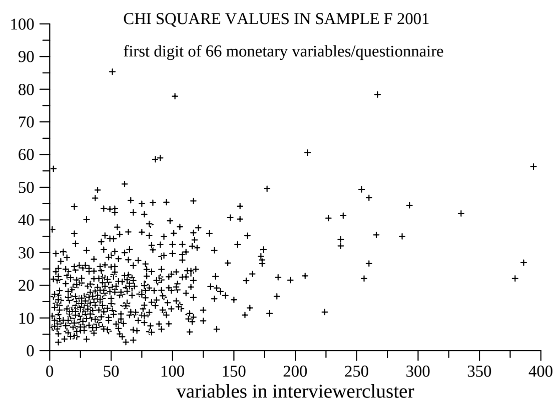


FIGURE 46: *First digit distribution: Chi-Square values for interviewer cluster in wave 1, sample F*

FIGURE 47: *First two digit distribution: Chi-Square values for interviewer cluster in wave 1, sample F*



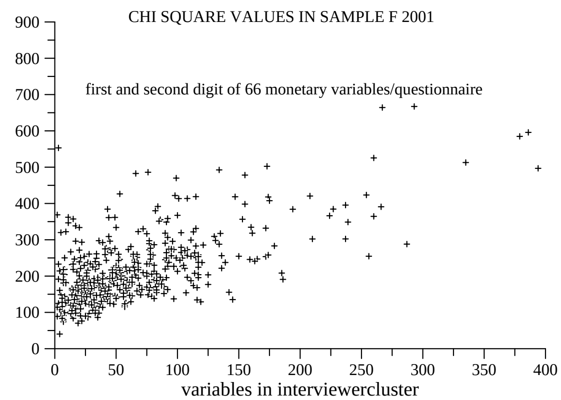
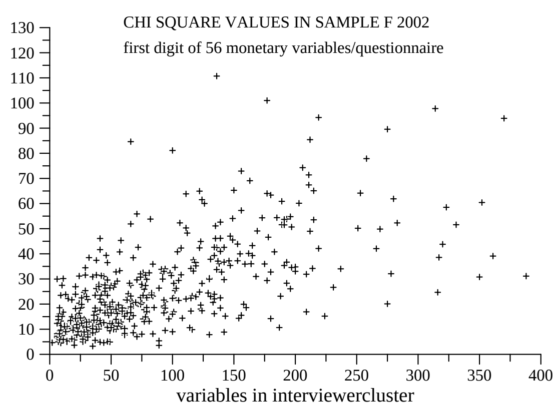


FIGURE 48: *First digit distribution: Chi-Square values for interviewer cluster in wave 2, sample F*

FIGURE 49: *First two digit distribution: Chi-Square values for interviewer cluster in wave 2, sample F*



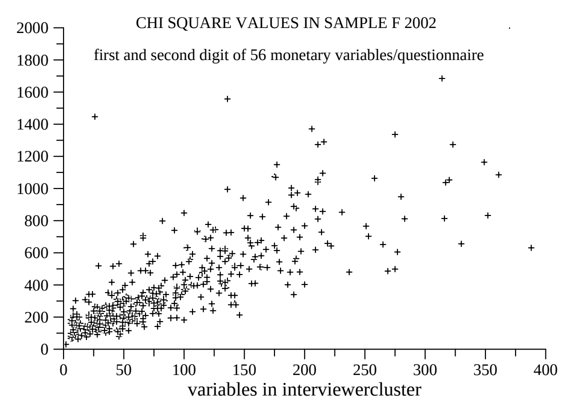


FIGURE 50: *First digit distribution: Chi-Square values for interviewer cluster in wave 3, sample F*

FIGURE 51: *First two digit distribution: Chi-Square values for interviewer cluster in wave 3, sample F*

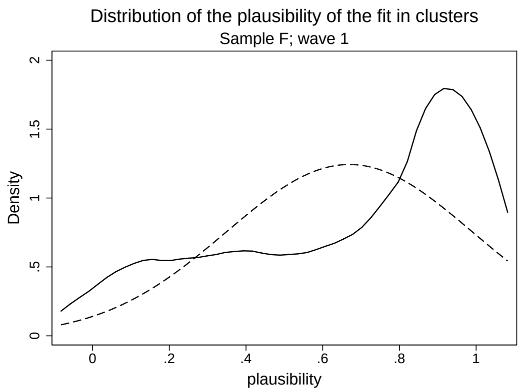


FIGURE 52: *Distribution of the probability $P(perc)$, sample F, wave 1, first digits*

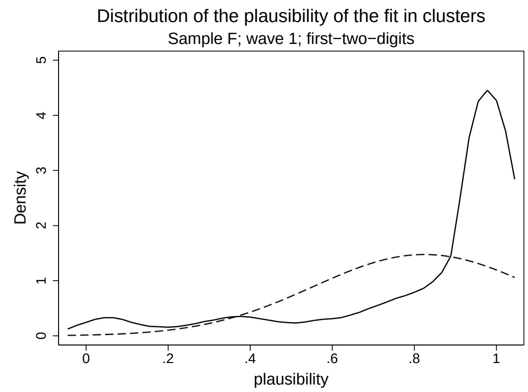


FIGURE 53: *Distribution of the probability $P(perc)$, sample F, wave 1, first two-digits*

TABLE 5: *Interviewer-ranking by plausibility of Interviewer clusters in wave 1-3, sample F, ($B = 2000$)*

		wave 1					wave 2					wave 3		
Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)
1	xx199x	3	55.23	0.0000	1	xx199x	3	55.23	0.0000	1	xx336x	253	64.239959	0
2	xx984x	9	74.85	0.0000	2	xx087x	39	49.17	0.0010	2	xx104x	314	97.889067	0
3	xx079x	136	77.41	0.0000	3	xx325x	51	85.38	0.0000	3	xx150x	211	71.466719	0
4	xx127x	124	73.43	0.0005	4	xx905x	37	46.70	0.0000	4	xx242x	275	89.652128	0
5	xx690x	124	59.11	0.0010	5	xx573x	267	78.39	0.0000	5	xx278x	136	110.74833	0
6	xx866x	101	44.66	0.0055	6	xx752x	44	43.49	0.0000	6	xx994x	411	130.98658	0
7	xx450x	231	80.53	0.0075	7	xx659x	20	44.03	0.0000	7	xx459x	211	67.486643	0
8	xx502x	116	48.25	0.0075	8	xx066x	30	40.16	0.0020	8	xx800x	258	77.954474	0
9	xx713x	32	46.09	0.0090	9	xx226x	102	77.87	0.0010	9	xx029x	413	108.8269	0
10	$\boxed{\textbf{xx085x}}$	**43**	**60.28**	**0.0120**	10	xx239x	42	33.33	0.0060	10	xx668x	206	74.326605	0
11	xx042x	124	45.39	0.0130	11	xx984x	5	29.70	0.0000	11	xx078x	219	94.223993	0
12	xx325x	60	45.49	0.0140	12	xx913x	20	35.78	0.0090	12	xx224x	100	81.085684	0
13	xx305x	128	44.89	0.0150	13	xx188x	7	25.17	0.0030	13	xx247x	66	84.643992	0
14	xx013x	108	44.65	0.0155	14	xx127x	3	21.89	0.0000	14	xx557x	6	29.953708	0
15	xx496x	48	54.40	0.0160	15	xx739x	86	58.56	0.0050	15	xx966x	177	101.03357	0
16	xx800x	183	65.23	0.0200	16	xx491x	45	35.24	0.0090	16	xx995x	212	85.439827	0
17	xx796x	102	43.54	0.0225	17	xx874x	210	60.60	0.0090	17	xx053x	486	113.5926	0
18	xx027x	24	43.41	0.0245	18	xx072x	13	24.95	0.0130	18	xx072x	26	136.96027	0
19	xx404x	20	32.25	0.0275	19	xx751x	30	30.69	0.0110	19	xx573x	464	251.24401	0
20	xx115x	92	42.01	0.0290	20	xx024x	21	32.75	0.0170	20	xx720x	370	93.718735	0
21	xx456x	19	33.32	0.0305	21	xx792x	11	30.27	0.0080	21	xx127x	215	65.142029	0.001
22	xx831x	74	50.87	0.0335	22	xx589x	11	30.03	0.0080	22	xx079x	180	63.390459	0.002
23	xx273x	22	40.29	0.0345	23	xx589x	61	51.01	0.0080	23	xx502x	189	60.928209	0.002
24	xx700x	23	39.38	0.0360	24	xx305x	9	27.33	0.0080	24	xx552x	203	60.096966	0.002
25	xx802x	31	33.10	0.0445	25	xx221x	7	22.68	0.0250	25	xx926x	156	72.969193	0.002
26	xx679x	23	34.65	0.0515	26	xx271x	14	28.48	0.0110	26	xx261x	444	84.122567	0.003
27	xx336x	138	48.98	0.0540	27	xx718x	5	24.14	0.0000	27	xx013x	163	69.12528	0.003
28	xx135x	87	37.71	0.0545	28	xx013x	90	58.99	0.0170	28	xx136x	122	64.968663	0.004
29	xx232x	138	48.28	0.0625	29	xx977x	44	30.91	0.0250	29	xx574x	280	61.966744	0.005
30	xx261x	108	37.03	0.0695	30	xx079x	24	26.13	0.0220	30	xx387x	11	30.136284	0.006
31	xx491x	78	44.07	0.0725	31	xx815x	49	43.29	0.0280	31	xx505x	196	54.849115	0.007
32	xx316x	18	28.16	0.0735	32	xx306x	27	25.27	0.0330	32	xx129x	193	53.902532	0.009
33	xx276x	27	33.88	0.0780	33	xx968x	15	24.13	0.0330	33	xx568x	177	64.057069	0.010
34	xx920x	107	35.43	0.0830	34	xx281x	53	43.47	0.0310	34	xx751x	215	53.604411	0.010
35	xx708x	106	35.58	0.0850	35	xx290x	32	25.20	0.0400	35	xx146x	126	60.060397	0.011
36	xx617x	53	36.45	0.0850	36	xx053x	6	21.49	0.0210	36	xx389x	124	61.575022	0.011
37	xx752x	48	41.87	0.0945	37	xx902x	4	17.25	0.0580	37	xx115x	191	53.746823	0.012
38	xx226x	116	33.77	0.0945	38	xx756x	13	20.49	0.0510	38	xx372x	111	63.806544	0.014
39	xx413x	62	31.04	0.1005	39	xx461x	53	42.30	0.0370	39	xx504x	251	50.254442	0.014
40	xx751x	93	33.58	0.1020	40	xx502x	84	45.22	0.0430	40	xx581x	191	51.574477	0.017
		⋮					⋮					⋮		
536	xx105x	4	12.51	1.0000	473	xx454x	114	5.72	1.0000	461	xx862x	54	13.928842	1.0000

Source SOEP, sample F, individual questionnaire, only monetary variables, 2000 - 2002 (own calculation)

5.7 Detecting unusual data rather than data not in conformity with Benford

In the previous sections we have assumed that Benford's Law holds completely for all our data sets. We have used Pearson's chi-square test statistic to determine whether an interviewer's data follow Benford's Law:

$$\chi_i^2 = n_i \sum_{d=1}^{9} \frac{(h_{d_i} - h_{b_d})^2}{h_{b_d}}$$

where n_i is the number of first digits in the interviewer cluster i, h_{d_i} is the observed proportion of digit $d = 1, \ldots, 9$ in interviewer cluster i and h_{b_d} is the proportion of digit d under Benford's distribution ($log_{10}(\frac{d+1}{d})$).

It may sometimes be useful to bear in mind the fact that Benford's Law may not hold completely for a particular data set. However, we can assume that the vast majority of interviewers are honest, meaning that the estimated value of h_{b_d} using the complete universe of data collected by all interviewers is close to the true value of h_{b_d} (cf. Swanson/Cho/Eltinge 2003). An alternative test statistic is therefore a chi-square statistic where we use instead of h_{b_d} the proportion of all numbers collected in survey whose leading digit is d, that is $\sum_{i=1}^{k} \frac{n_{d_i}}{\sum_{d=1}^{9} n_{d_i}}$, hence we get the formula:

$$\chi_i^2 = n_i \sum_{d=1}^{9} \frac{(h_{d_i} - \sum_{i=1}^{k} \frac{n_{d_i}}{\sum_{d=1}^{9} n_{d_i}})^2}{\sum_{i=1}^{k} \frac{n_{d_i}}{\sum_{d=1}^{9} n_{d_i}}} \tag{9}$$

n_i is the number of first digits in the interviewer cluster $i = 1, \ldots, k$, h_{d_i} is the observed proportion of digit $d = 1, \ldots, 9$ in interviewer cluster i and n_{d_i} is the number of digit d in cluster i.

Table 6 on page 34 shows the interviewer ranking by the plausibility of the obtained chi-square value based on equation 9. Again, the known fabricated clusters are framed and marked. We can see that for sample A/B, the first cheating interviewer is ranked at position 16 followed by the second at position 20. If we compare this ranking with the list in table 2 on page 24, which is based on the assumption that the data set used follows Benford exactly, we find no evidence that the modified test statistic yields better results. The first cheating interviewer is ranked at position 1 in the Benford ranking list and the next falsifier later on at rank 61. For sample E, the Benford assumption yields clearly better results. Table 4 on page 29 shows that three falsifiers can be found among the top seven, whereas in the list based on the modified test statistic in table 6 only two cheating interviewers are in the top ten. The unusual pattern method ranks the detected fake at position no. 7 instead of no. 10 in the case of Benford in sample F. However, the new falsifier detected by Benford in sample F could not be identified with the unusual pattern method (no. 52).

Our results suggest that the calculation of both test statistics could be useful. If we assume that the first twenty interviewers in the ranking list can be classified as suspect, we get quite similar suspicious interviewer clusters. Furthermore, under this criterion we find two falsifiers in sample A/B with the modified test statistic as opposed to only one with Benford.

However, if we use two different test statistics and get two ranking lists, the question arises as to whether we have the same suspect interviewers at the top of our lists. We would expect a positive correlation of both rankings. The figures 54-56 on pages 35 and 36 show Spearman's correlation coefficient for the rankings based on Benford and the modified test statistic by the size of the sorted list. We can assume that the correlation differs depending on whether we use only the top twenty or the whole list. We sort both combined ranking lists by Benford and by the ranks of the unusual data statistic. The graphs show particularly high correlations for interviewers at the top. For the top ten we have values of 0.7 (sample A) and nearly 1.0 (sample E). The correlation swings into a value of 0.5 if we enlarge the number of clusters included, and then increases slightly with the number of clusters. This finding suggests in particular that the clusters with the worst plausibility in both lists are highly positively correlated. Both statistics therefore tend to classify the same interviewer as suspect.

6 Explaining the deviation from Benford's distribution

Thus far we have used Benford's Law to help to identify sources of unusual data. We have recognized the data from each interviewer as arising from a simple random sample and used Pearson's chi-square test statistic to determine whether the field representative's data follow Benford's Law. At this point, the question arises as to why, in several cases, non-faked clusters do not conform to Benford's Law. In the following section, we will try to develop and test hypotheses about factors that may affect our test statistic and use a regression framework to give an empirical explanation.

TABLE 6: *Detecting unusual data in interviewer clusters in sample A/B, sample E, and sample F, wave 1 - ranking by plausibility*

sample A/B, wave 1					sample E, wave 1					sample F, wave 1				
Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)	Rank	Intnr	digits	chi-sq.	P(perc)
1	xx654x	226	61.51	0.0000	1	xx492x	4	29.26	0.0000	1	xx491x	78	66.19	0.0000
2	xx111x	713	67.39	0.0000	2	$\boxed{\textbf{xx928x}}$	**7**	**52.06**	**0.0016**	2	xx199x	3	77.37	0.0000
3	xx147x	94	58.81	0.0000	3	xx550x	27	41.56	0.0023	3	xx057x	75	68.51	0.0000
4	xx552x	3	36.34	0.0000	4	xx553x	105	38.77	0.0648	4	xx305x	128	49.85	0.0000
5	xx687x	27	31.70	0.0010	5	xx232x	61	40.90	0.0661	5	xx079x	136	51.65	0.0000
6	xx756x	58	37.06	0.0020	6	xx066x	40	35.85	0.0727	6	xx127x	124	47.34	0.0010
7	xx106x	7	35.87	0.0040	7	xx843x	95	36.69	0.0907	**7**	$\boxed{\textbf{xx085x}}$	**43**	**98.28**	**0.0010**
8	xx320x	16	32.25	0.0050	8	$\boxed{\textbf{xx928x}}$	**158**	**33.67**	**0.1271**	8	xx690x	124	39.71	0.0020
9	xx512x	90	30.31	0.0050	9	xx708x	136	31.26	0.1313	9	xx013x	108	40.16	0.0020
10	xx887x	29	28.09	0.0090	10	xx690x	27	22.33	0.1851	10	xx325x	60	65.60	0.0030
11	xx003x	32	22.20	0.0100	11	xx029x	14	20.37	0.2133	11	xx505x	157	51.77	0.0050
12	xx401x	26	25.28	0.0170	12	xx905x	18	19.45	0.2523	12	xx502x	116	36.50	0.0070
13	xx583x	20	30.28	0.0190	13	xx160x	127	28.74	0.2735	13	xx857x	114	33.64	0.0100
14	xx778x	137	28.26	0.0210	14	xx469x	173	28.34	0.3046	14	xx984x	9	57.91	0.0110
15	xx752x	24	25.62	0.0230	15	xx693x	9	21.94	0.3106	15	xx302x	106	37.49	0.0150
16	$\boxed{\textbf{xx800x}}$	**91**	**21.72**	**0.0240**	16	xx500x	27	18.86	0.3331	16	xx116x	62	44.91	0.0160
17	xx544x	76	25.18	0.0240	17	xx668x	123	26.43	0.3429	17	xx708x	106	33.90	0.0190
18	xx342x	94	24.61	0.0250	18	xx690x	137	22.54	0.3980	18	xx062x	122	29.54	0.0230
19	xx676x	170	28.89	0.0300	19	$\boxed{\textbf{xx683x}}$	**221**	**27.46**	**0.4322**	19	xx336x	138	37.83	0.0280
20	$\boxed{\textbf{xx827x}}$	**122**	**29.11**	**0.0310**	20	xx234x	95	21.70	0.4435	20	xx871x	43	61.78	0.0280
21	xx192x	32	18.56	0.0320	21	xx990x	103	23.07	0.4639	21	xx450x	231	41.20	0.0290
22	xx998x	73	24.30	0.0370	22	xx761x	143	20.56	0.4646	22	xx438x	106	30.91	0.0290
23	xx650x	32	17.41	0.0460	23	xx242x	13	16.17	0.4825	23	xx700x	23	71.79	0.0290
24	xx353x	4	24.19	0.0470	24	xx607x	27	16.13	0.4934	24	xx404x	20	40.60	0.0340
25	xx950x	27	20.45	0.0470	25	xx530x	33	15.94	0.5080	25	xx027x	24	56.28	0.0370
26	xx785x	28	19.90	0.0480	26	xx278x	111	20.68	0.5201	26	xx042x	124	27.91	0.0370
27	xx020x	7	25.00	0.0520	27	xx020x	67	21.00	0.5235	27	xx154x	57	41.74	0.0380
28	xx234x	89	19.76	0.0540	28	xx674x	85	18.40	0.5317	28	xx720x	97	31.58	0.0390
29	xx277x	23	20.45	0.0610	29	xx262x	24	18.27	0.5485	29	xx372x	61	40.59	0.0390
30	xx884x	7	23.14	0.0610	30	xx037x	89	17.78	0.5580	30	xx800x	183	38.98	0.0410
31	xx216x	45	21.35	0.0620	31	xx716x	76	20.77	0.5606	31	xx731x	89	33.62	0.0440
32	xx365x	32	16.75	0.0640	32	xx641x	94	18.89	0.5756	32	xx192x	128	27.05	0.0450
33	xx582x	170	25.37	0.0650	33	xx336x	32	13.70	0.5901	33	xx866x	101	26.81	0.0490
34	xx866x	61	23.14	0.0680	34	xx589x	109	19.51	0.6076	34	xx245x	74	42.25	0.0490
35	xx440x	21	19.82	0.0700	35	xx134x	103	19.66	0.6203	35	xx956x	96	29.53	0.0540
36	xx167x	90	17.64	0.0750	36	xx686x	21	13.94	0.6351	36	xx864x	10	34.99	0.0540
37	xx508x	37	21.16	0.0760	37	xx256x	18	12.63	0.6370	37	xx476x	91	30.24	0.0580
38	xx395x	18	21.58	0.0780	38	xx069x	43	12.73	0.6379	38	xx203x	55	39.59	0.0590
39	xx275x	25	20.32	0.0810	39	xx051x	107	18.90	0.6400	39	xx047x	104	27.57	0.0660
40	xx352x	44	21.92	0.0890	40	xx683x	75	19.06	0.6679	40	xx787x	131	25.60	0.0690
		⋮					⋮					⋮		
636	xx169x	18	2.74	1.0000	150	xx239x	6	11.60	1.0000	536	xx163x	79	6.16	1.0000

Source SOEP, sample A/B, sample E and sample F, individual questionnaire, only continuous variables (own calculation)

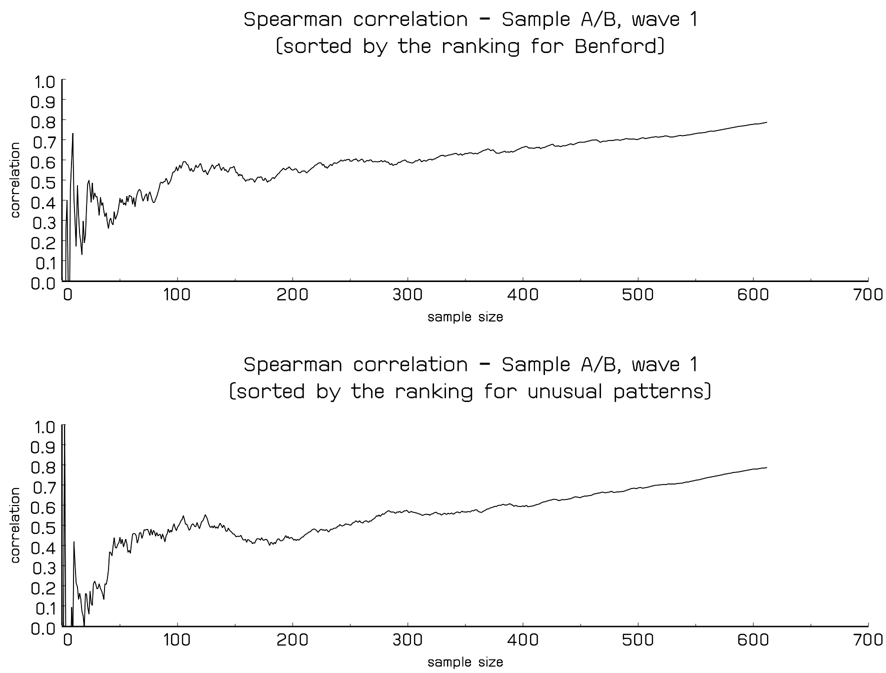


FIGURE 54: *Spearman's correlation for the Benford and the 'unusual pattern' ranking (sample A/B, wave 1)*

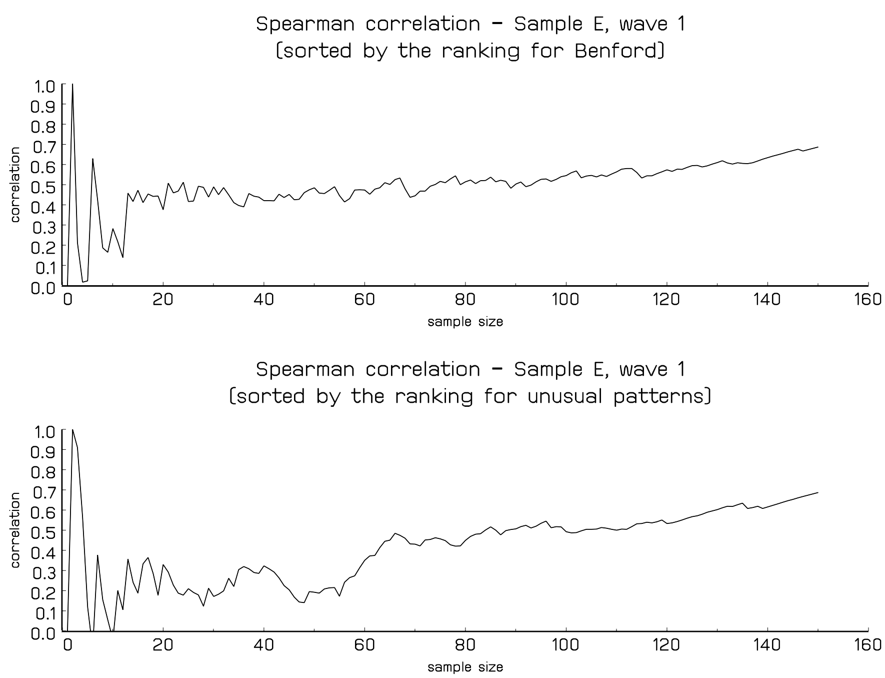


FIGURE 55: *Spearman's correlation for the Benford and the 'unusual pattern' ranking (sample E, wave 1)*

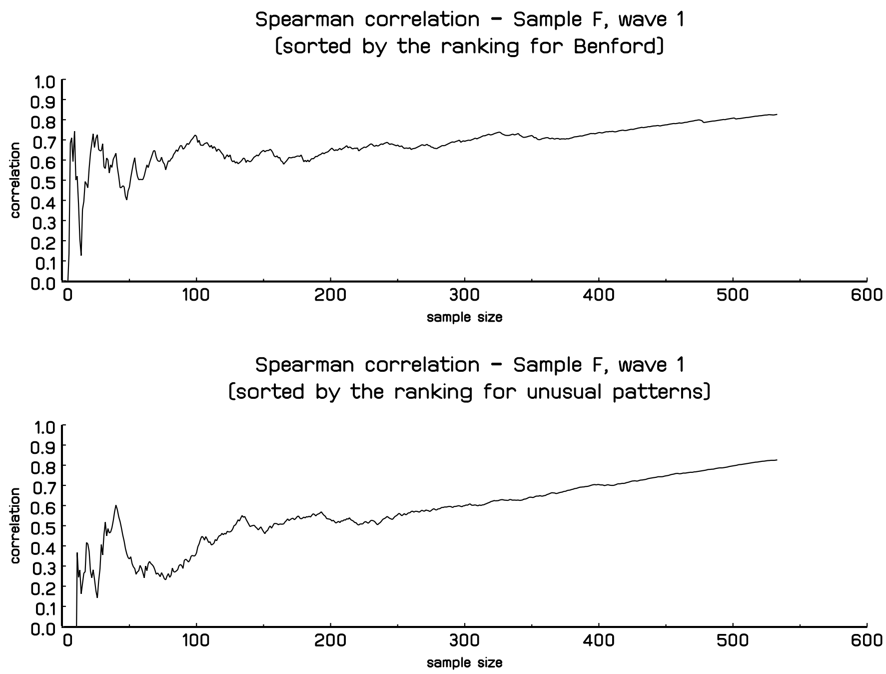


FIGURE 56: *Spearman's correlation for the Benford and the 'unusual pattern' ranking (sample F, wave 1)*

6.1 Hypotheses

6.1.1 Homogeneity

We can assume that homogeneity is one important reason why Benford's Law does not hold in particular clusters.

Homogeneity in data can be due to two reasons: 1. The interviewer has fabricated the values and has used quite similar values. 2. The interviewer works in a very homogeneous area. It may be that the distribution of the continuous values in an interviewer cluster are quite similar because the field representative works in an area where respondents tend to have similar living conditions and incomes. If the first digit distribution for each respondent always deviates in the same way from the logarithmic distribution within a particular cluster, we can expect an accumulation and a higher disproportion of certain digits. The test statistics, therefore, show higher values for these homogeneous clusters.

Unfortunately, in the SOEP, the interviewers are not randomly assigned to areas like most other big household surveys[19]. We cannot, therefore, distinguish between these two reasons.

We can assume that in the case of a homogeneous area, the respondents tend to have close gross income values, or that the standard deviation of the gross income variable is lower than in heterogeneous areas. If we consider the chi-square values for sample E, 1998, in scatterplot figure 57 (left-hand side) on page 37 we can observe, very close to the two faked clusters, an (assumed) valid cluster (circled) with a chi-square value of 43.47 in the case of $n = 177$ digits. The distribution of the standard deviations by the mean of income is shown on the right-hand side of figure 57.[20]

[19] An exception is the experimental subsample in the BHPS, wave 2. Nevertheless, this experiment was conducted for only a quarter of the full sample in sample 2.

[20] We calculate the standard deviation of the mean values only for clusters with at least three gross income values. Three small falsified clusters have only two gross income values and they were excluded from the scatterplot.

We can see that in comparison with the other clusters, the circled cluster has a very low standard deviation and that this cluster seems to be rather homogeneous.

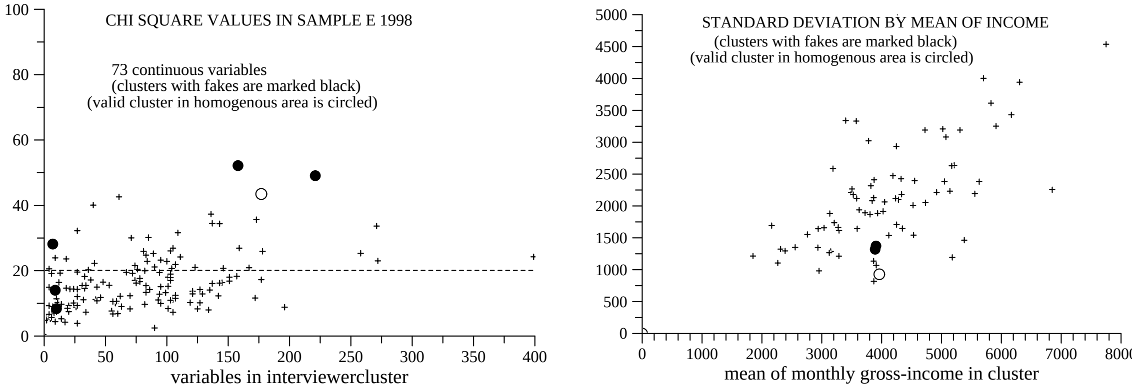


FIGURE 57: *Chi-square values (first digit distr.) and standard deviation of gross-income in sample E, individual questionnaire, 1998*

To compare interviewer clusters, in our analysis, we will use the variation coefficient vc_j of the gross income x_{ij} (see equation 10). This can be calculated as the ratio of the standard deviation and the mean of the gross income $\bar{x_j}$ in each cluster j.

$$vc_j = \frac{\sqrt{\frac{1}{n}(x_{ij} - \bar{x_j})^2}}{\bar{x_j}} \tag{10}$$

The lower the values of vc_j, the lower the standard deviation of the gross-income compared to the mean in each cluster. Table 7 shows some statistics of the variation coefficient for each subsample. We can see that in sample C (East German sample) the average mean of the variation in each cluster is lowest and in sample F highest in the first three waves.

6.1.2 Rounding

Rounding behavior is not unusual in surveys. Normally a respondent is not able to recall the exact value of his monthly income without looking at his wage slip. Based on this, some slight rounding behavior would be acceptable in an interview situation, but strong rounding could cause a loss of information and may even lead to incorrect conclusions of empirical analyses. Empirical studies show that rounding income values depends on interviewer and respondent characteristics, as well as on the data collection method used. Schräpler (1999) showed a decrease in rounding depending on the age of the interviewer and the duration of the interview. Furthermore, rounding is lower in self- completed questionnaires as compared to face-to-face interviews and increases with income level in the SOEP.

It is reasonable to assume that rounding of monetary values causes bad fits in the case of the first two digit distribution. We have seen in section 4.1.3 of this chapter that values like $10, 20, 30, \ldots, 90$ have higher proportions than predicted in the case of the first two digit distributions. We can therefore expect that interviewer clusters with many rounded values will have an inferior Benford fit for the first two digit distribution as compared to clusters with fewer rounded values.

In order to check the impact of rounding it is necessary to specify a meaningful statistic. We can assume that the higher the proportion of zeros on all significant digits of a monetary value, the stronger the rounding behavior (without decimal places).[21] Following this idea, in equation 11, we use a rounding coefficient rd_j that is the average proportion of the number of zeros n_{0i} on

[21]For example the value '3,000' has a proportion of 3/4 and the value '3,200' only 2/4.

TABLE 7: *Gross income homogeneity measured by variation coefficient in each subsample*

	Obs.	Mean	Std. Dev.	Min	Max
Sample A/B					
wave 1	557	0.513	0.220	0.000	1.350
wave 2	401	0.504	0.223	0.000	1.754
wave 3	351	0.515	0.208	0.023	1.557
Sample C					
wave 1	209	0.365	0.149	0.076	1.139
wave 2	237	0.393	0.161	0.000	0.922
wave 3	223	0.393	0.177	0.012	1.110
Sample E					
wave 1	108	0.518	0.226	0.025	1.277
wave 2	96	0.553	0.252	0.017	1.463
wave 3	103	0.577	0.259	0.018	1.098
Sample F					
wave 1	467	0.581	0.242	0.000	1.437
wave 2	404	0.596	0.240	0.000	1.426
wave 3	384	0.613	0.255	0.015	1.663

Source: SOEP samples A/B, C, E, and F, individual questionnaires (own calculation)

the number of all digits of an entire value n_{di}, calculated for values $i = 1, \ldots, n_j$ in an interviewer cluster j. The higher the value of this coefficient, the stronger the rounding behavior in the cluster is.

$$rd_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{n_{0i}}{n_{di}} \tag{11}$$

Table 8 shows the mean and standard deviation as well as the minimum and maximum values in the first three waves of each subsample analyzed. Samples A/B and E have quite similar mean values whereas the means in sample C are slightly lower and in sample F, slightly higher.

TABLE 8: *Descriptive statistics of the rounding coefficient*

	Obs	Mean	Std. Dev.	Min	Max
Sample A/B					
wave 1	636	0.465	0.118	0.000	0.800
wave 2	463	0.284	0.086	0.000	0.569
wave 3	410	0.469	0.092	0.000	0.722
Sample C					
wave 1	214	0.365	0.063	0.000	0.546
wave 2	264	0.376	0.110	0.000	0.750
wave 3	278	0.442	0.087	0.184	0.750
Sample E					
wave 1	150	0.471	0.110	0.000	0.767
wave 2	125	0.453	0.083	0.167	0.667
wave 3	129	0.469	0.083	0.149	0.722
Sample F					
wave 1	536	0.480	0.081	0.000	0.667
wave 2	473	0.480	0.082	0.000	0.775
wave 3	461	0.483	0.076	0.167	0.692

Source: SOEP samples A/B, C, E, and F, individual questionnaires (own calculation)

6.1.3 Data collection methods

Data collection methods may have an impact on a respondent's as well as on an interviewer's behavior. The SOEP is a multimethod survey. Table 9 shows the average proportion of the interview modes used in the clusters of the analyzed subsamples for the first three waves. Sample A/B contains West German respondents (sample A) and foreigner respondents who live in West Germany (sample B). Sample A is mainly administered by face-to-face (face) and partly by self-completed interviews (in the presence of the interviewer). In a small number of interviews, a mixed mode (mix) is used. In foreigner sample B, we distinguish between interviews with (mdolm) and without (odolm) an additional interpreter. A small proportion is carried out by telephone or mail if an interview would, otherwise, not be possible (not shown in the table). Note that we do not analyze these interviews because they are not conducted by particular interviewers.

Sample C contains only East German respondents. In wave 1 all interviews were carried out by interviewers and personal interviewing, but there is no information about the method used. Waves 2 and 3 were conducted mainly face-to-face. Like the other subsamples, the extension samples E and F were carried out using face-to-face and self-completion modes as well as using CAPI (computer assisted personal interviewing).

Apart from the CAPI interviews in sample E, wave 1, which are part of an experimental setting in the SOEP, the decision about the data collection mode used is not predetermined by the fieldwork organization. The decision is reached based on the interaction between the interviewer and the respondent in the interview situation. Past experiences show that, in some interview situations, the self-completed mode in the presence of the interviewer might be more practical than the face-to-face mode. There is also some empirical evidence that the self-completion mode results in more accurate answers than orally given answers in the SOEP (Schräpler 1999).

TABLE 9: *Average proportion of data collection methods in the clusters (in percent)*

	face	self	mix	dolm	odolm		n
Sample A/B							
wave 1	49.65	21.60	8.14	4.47	16.06	100	631
wave 2	47.75	18.67	11.18	4.36	18.04	100	461
wave 3	48.90	20.60	8.82	3.98	17.70	100	408
Sample C							
wave 1	n.k.	n.k.	n.k.	-	-	-	214
wave 2	56.29	21.27	22.43	-	-	100	264
wave 3	54.56	25.83	19.60	-	-	100	277
Sample E				capi			
wave 1	45.16	10.60	10.25	33.99		100	150
wave 2	37.13	18.36	4.89	39.56		100	125
wave 3	18.58	17.22	5.90	57.44		100	129
Sample F							
wave 1	40.36	21.53	8.64	29.48		100	536
wave 2	34.76	24.71	6.49	34.02		100	473
wave 3	33.40	25.52	7.51	34.48		100	461

Source: SOEP sample A/B, C, E, and F, individual questionnaires (own calculation)

6.2 Modeling the fit to Benford

Next we model the fit to Benford by using a linear regression framework with the chi-square values of the first and first two digit distribution as the continuous dependent variable. Because of the skewness of the chi-square distributions we transform the original values y_{it} to normality using a Box-Cox transformation

$$y_{it}^* = \begin{cases} \frac{y_{it}^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log(y_{it}), & \text{if } \lambda = 0 \end{cases}$$

where λ could be found by maximum likelihood. Figure 58 shows the distribution of the original and transformed chi-square values for the waves 1, 2, and 3 of sample A/B and the normal density distribution (dashed line). We can see the positive skew of the original distribution. After the Box-Cox transformation using $\lambda = -0.133$ we get nearly a normality distribution with a skewness of < 0.001. Table 10 shows for all chi-square distributions the estimated values of parameter λ and

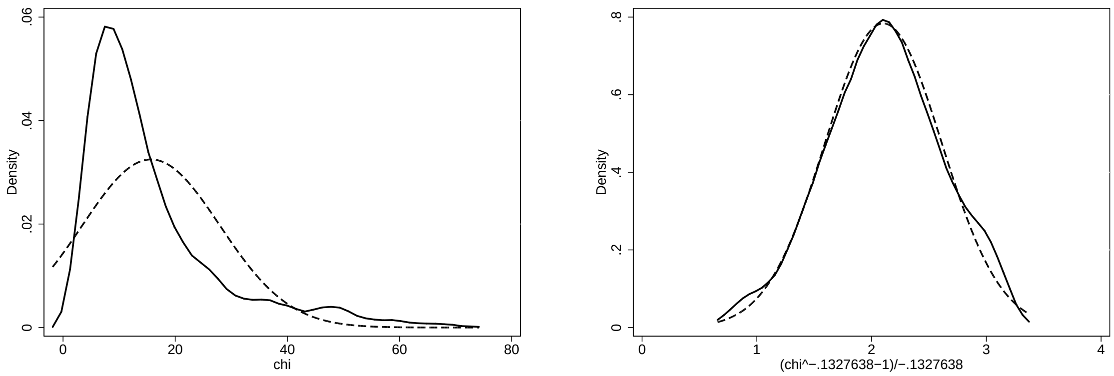


FIGURE 58: *Distribution of the original (chi) and transformed chi-square values for the waves 1,2 and 3 of sample A/B (dashed line indicates normal density)*

the remaining skewness.[22] We use the transformation $log(y_{it})$ for sample C and sample F (here only for the first digit chi-square distribution) because the estimated values of parameter λ are in these cases almost 0.

TABLE 10: *Box-Cox transformation parameter λ for all chi-square distributions*

	first digit		*first two digits*	
	λ	skewness	λ	skewness
Sample A/B	-0.1328	-0.0001	-0.1290	-0.0001
Sample C	-0.0038	0.0000	0.0386	0.0003
Sample E	0.1258	0.0000	0.3214	-0.0002
Sample F	-0.0010	0.0000	-0.2731	-0.0008

Then we can write the following equation

$$y^*_{it} = \alpha + x'_{it}\beta + \epsilon_{it} \tag{12}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$. α is a scalar, β is a vector $K \times 1$ and x_{it} is the ith observation on k explanatory variables.

To control for individual-specific effects we use panel data models, either a fixed effects or a random effects model. To do so, we have to specify a complex error component:

$$w_{it} = u_i + e_{it}$$

where u_i are cross-section specific components and ϵ_{it} are remainder effects. The following passage gives a short description of these panel data models; more details can be found in Green (2003) and Hsiao (1986).

[22]The estimation is done with STATA and the procedure *bcskew0*.

6.2.1 Fixed effects model

If the cross-section-specific components u_i are thought of as fixed parameters, we have to estimate α_i $i = 2, \ldots, N$ individual effects, that are specific for each respondent but constant over time. The model lets us use the changes in the variables over time to estimate effects of the independent variables on our dependent variable.

$$y^*_{it} = \alpha + x'_{it}\beta + \sum_{i=2}^{n} \alpha_i D_i + \epsilon_{it} \tag{13}$$

where D_i is a dummy variable for the ith individual. Using OLS on equation 13 leads to the least squares dummy variable (LSDV) estimator. If equation 13 is the true model, LSDV is BLUE (best linear unbiased estimator) as long as ϵ_{it} is the standard *i.i.d.* (independent identically distributed) disturbance with mean 0 and variance matrix $\sigma^2_\epsilon I_{nT}$.

6.2.2 Random effects model

The fixed effects model is appropriate when differences between individual agents (here the interviewer) may reasonably be viewed simply as parametric shifts in the regression itself. The disadvantage of the fixed effect model is that there may be many parameters. The loss of freedom can be avoided if the term u_i can be assumed as random. Assume $u_i \sim i.i.d\ (0, \sigma^2_u)$ and $\epsilon_{it} \sim i.i.d\ (0, \sigma^2_\epsilon)$, and the u_i are independent of ϵ_{it}. In addition, the explanatory variables X_{it} are independent of the u_i and ϵ_{it} for all i and t. The specification of random effect models implies a homoskedastic variance $Var(w_{it}) = \sigma^2_u + \sigma^2_\epsilon$ for all i and t, and serial correlation over time only between the disturbances of the same individual.

$$Cov(w_{it}, w_{js}) = \sigma^2_u + \sigma^2_\epsilon \quad \text{for} \quad i = j, t = s \tag{14}$$

$$= \sigma^2_u \quad \text{for} \quad i = j, t \neq s \tag{15}$$

and zero otherwise. This also means that the correlation coefficient between w_{it} and w_{js} is

$$\rho = Cov(w_{it}, w_{js}) = 1 \quad \text{for} \quad i = j, t = s \tag{16}$$

$$= \sigma^2_u/(\sigma^2_u + \sigma^2_\epsilon) \quad \text{for} \quad i = j, t \neq s \tag{17}$$

and zero otherwise. Under the random effects model, GLS based on the true variance component is BLUE, and the feasible GLS estimator are asymptotically efficient as either n or $T \rightarrow \infty$.

Testing for random effects Heteroskedasticity occurs when the assumption that residual variance is constant across all observations in the data set is violated. The OLS estimates of coefficients remains unbiased but it can be shown that the OLS estimates of the standard errors (and hence t and F tests) are biased. If heteroskedasticity is present, we should use a more efficient method such as GLS instead of OLS. Breusch and Pagan (1979) derived a Lagrange multiplier (LM) test for the random effects model based on the OLS residuals. The specific hypothesis under investigation is the following:

$$H_0 \ : \ \sigma^2_u = 0 \quad (\text{or} \quad Corr(w_{it}, w_{js}) = 0 \quad \text{for} \quad i = j \tag{18}$$

$$H_A \ : \ \sigma^2_u \neq 0 \tag{19}$$

If H_0 is rejected, we have to assume that heteroskedasticity is present. The Breusch-Pagan test statistic implemented in STATA (StataCorp LP 2003) is as follows:

$$LM = \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^{n} (\sum_t = 1^T e_{it})^2}{\sum_{i=1}^{n} \sum_{t=1}^{T} e^2_{it}} - 1 \right] \sim \chi^2_1$$

6.2.3 Hausman Test for Fixed or Random Effects

The generally accepted way of choosing between fixed and random effects is by running a Hausman test. Hausman (1978) derived a test based on the idea that according to the hypothesis of no correlation, both OLS in the LSDV model and GLS are consistent, but OLS is inefficient. The Hausman test checks a more efficient model against a less efficient but consistent model to make sure that the more efficient model also gives consistent results.

According to the alternative, OLS is consistent, but GLS is not. According to the null hypothesis, therefore, the two estimates should not differ systematically, and a test can be based on the difference. To test the difference, we need the covariance matrix of the difference vector $(b - \hat{\beta})$, where b is the OLS in LSDV, and $\hat{\beta}$ is GLS.

$$Var(b - \hat{\beta}) = Var(b) + Var(\hat{\beta}) - Cov(b, \hat{\beta}) - Cov(b, \hat{\beta})^{'}$$

Hausman's key result is that the covariance of an efficient estimator with its difference from an efficient estimator is zero, which implies that

$$Cov[(b - \hat{\beta}), \hat{\beta}] = Cov(b, \hat{\beta}) - Var(\hat{\beta}) = 0$$

or

$$Cov(b, \hat{\beta}) = Var(\hat{\beta})$$

Denote

$$Var(b - \hat{\beta}) = Var(b) - Var(\hat{\beta}) = \Sigma$$

The chi-squared test is based on the Wald criterion:

$$W = \chi^2_k = (b - \hat{\beta})^{'} \hat{\Sigma}^{-1} (b - \hat{\beta})$$

For $\hat{\Sigma}^{-1}$, we use the estimated covariance matrices of the slope estimator in the LSDV model and the estimated covariance matrix in the random effects model, excluding the constant term.

Hence, the Hausman method tests the null hypothesis that the coefficients estimated by the efficient random effects estimator are the same as the ones estimated by the consistent fixed effects estimator. If $Prob > \chi^2_k$ is larger than .05 then it is assumed that it is safe to use random effects. In the case of a significant P-value, we should use fixed effects.

6.3 Estimates

The tables 11, 12, 13, and 14 show the estimates of the linear panel models for the subsamples A/B, C, E, and F of the SOEP. We use the first three waves and estimate two models for each subsample.[23] In the first model we specify as the dependent variable the transformed chi-square value of the first digit distribution and in the second model the transformed chi-square value of the first two digit distribution. In all cases we perform a Breusch-Pagan test to check whether unobserved heterogeneity is present. The results show (test section at the bottom of the tables) that we have to take unobserved heterogeneity into account. The probability $Prob > \chi^2_1$ is, in all cases, lower than 0.05 and H_0 (homoskedasticity) has to be rejected. In addition we use a Hausman test (Hausman 1978) to examine if a fixed effects model or a random effects model is appropriate. The results suggest in five cases that a random effects model ($Prob > \chi^2_k$ is higher 0.05) and in three cases a fixed effects model.

The largest subsample, A/B, shown in table 11 contains a total of $N = 1,291$ observations from 579 interviewer clusters. Subsample F (table 14) has $1,254$ observations from 523 interviewers;

[23]The estimation is done with Stata 8.0 (StataCorp LP 2003).

Table 11: *Regression models for the transformed chi-square values of clusters in sample A/B, waves 1-3*

	first digit			*first two digit*		
	Random Effects Model			Fixed Effects Model		
variable	Coef.	z	$P > z$	Coef.	z	$P > z$
intercept	2.026	20.01	0.000	3.333	44.26	0.000
w2	0.503	11.70	0.000	0.413	14.57	0.000
w3	0.056	1.72	0.086	0.032	1.66	0.097
workload	0.009	17.49	0.000	0.008	15.26	0.000
face	0.000	0.28	0.782	-0.001	-1.89	0.059
mbgl	0.001	0.88	0.379	0.001	0.96	0.336
vc	-0.147	-2.96	0.003	0.006	0.16	0.872
round	-0.244	-1.53	0.125	0.432	3.67	0.000
refuse	-0.011	-1.55	0.121	0.000	0.02	0.981
HH-contacts	-0.013	-1.18	0.238	-0.014	-1.87	0.063
σ_u	0.146			0.147		
σ_e	0.342			0.150		
ρ	0.153			0.000		
R^2	0.455			0.488		
N	1291			1291		
interviewer	579			579		
Box-Cox λ	-0.133			-0.129		
Breusch/Pagan						
χ^2_1	14.20	($p = 0.000$)		38.47	($p = 0.000$)	
Hausman						
χ^2_9	13.09	($p = 0.159$)		17.48	($p = 0.042$)	

Source: SOEP, sample A/B, wave 1-3 (own calculation)

subsamples C and E are distinctly smaller. The overall fit of the linear models is indicated by the explained variance R^2. The explained variance ranges from 0.09 to 0.61 and is always higher for the first two digit regression.

The estimates in all subsamples indicate for the first and the first two digit distributions significant increasing chi-squared values caused by increasing workloads. This was to be expected because higher workloads indicate more digits in the clusters.

However, we are more interested in the effects of homogeneity, rounding, and data collection methods. Homogeneity is measured by the variation coefficient *vc*. We assume that, with higher values of *vc*, the fit to Benford's Law will be enhanced. The estimates show in all subsamples (except in sample E) significant negative coefficients for the first digit distribution. This supports our first hypothesis and means that the lower the variability of income in the clusters, the higher the chi-square values. In spite of this, in the case of the first two digit distribution, we find inconsistent results for samples E and F. They suggest a positive relationship.

In our second hypothesis, we state that rounding of continuous values will cause bad fits for the first two digit distribution. The estimates confirm this assumption: rounding is significant in all samples and by far the highest positive coefficient. For the first two digit distribution, an increase of *round* causes a strong increase in the transformed chi-square values. The coefficient of *round* is highest in subsample E and lowest in subsample A/B.

An interesting question is whether we can find a data collection effect. We can assume that the data collection method has an effect on a respondent's and an interviewer's behavior and on the way in which a respondent's answers are stored. However, we only find significant negative effects in sample E for the CAPI mode.

In addition, we control for the average number of refusals and unusable values as well as for the average number of necessary household contacts in the cluster. Descriptive statistics of refusals and HH contacts can be found in table 18 and 17 in the appendix.

We have shown in Schräpler (2004) that in sample A/B, wave 1, cheating interviewers underestimate the number of refusals (or don't knows) of the respondents in their questionnaires. If the assumption holds that suspicious interviewers have lower missing value rates and worse fit values, we would expect negative coefficients in our models. We can see that the estimated coefficients are in most cases negative but not significant. Hence, the results do not support our assumption.

The variable 'household contacts' (HH-contact) measures the influence of an increase in the average number of necessary contacts, to achieve the interviews, on the transformed chi-square values.[24] We find significant negative coefficients for samples A/B and F and positive coefficients in sample E. These results are therefore ambiguous and hard to interpret. On the one hand, it may be that suspicious interviewers need fewer contacts because they never enter the household and give only estimated values. On the other hand, interviewers with a higher workload may perform better, act more professionally, and need fewer contacts than interviewers with only a few interviews.

TABLE 12: *Regression models for the transformed chi-square values of clusters in sample C, wave 1-3*

variable	*first digit* Fixed Effects Model Coef.	z	$P > z$	*first two digit* Random Effects Model Coef.	z	$P > z$
intercept	3.494	18.40	0.000	4.801	65.81	0.000
w2	-0.780	-15.00	0.000	-0.773	-29.24	0.000
w3	-0.585	-10.42	0.000	-0.314	-11.21	0.000
workload	0.035	7.71	0.000	0.016	14.94	0.000
vc	-1.072	-6.24	0.000	-0.123	-1.77	0.077
round	-0.666	-1.64	0.102	0.896	5.66	0.000
refuse	-0.005	-0.38	0.701	-0.013	-2.32	0.020
unusable	-0.283	-1.54	0.124	-0.091	-1.34	0.182
σ_u	0.390			0.147		
σ_e	0.419			0.251		
ρ	0.464			0.255		
R^2	0.481			0.612		
N	669			669		
interviewer	324			324		
Breusch/Pagan						
χ^2_1	10.42	($p = 0.001$)		21.3	($p = 0.000$)	
Hausman						
χ^2_9	14.93	($p = 0.037$)		4.27	($p = 0.748$)	

Source: SOEP, sample C, wave 1-3 (own calculation)

[24]The variable 'household contacts' is not available in the first waves of sample C.

Table 13: *Regression models for the transformed chi-square values of clusters in sample E, wave 1-3*

	first digit			*first two digit*		
	Random Effects Model			Random Effects Model		
variable	Coef.	z	$P > z$	Coef.	z	$P > z$
intercept	3.190	7.57	0.000	8.865	10.27	0.000
w2	0.047	0.37	0.714	0.100	0.39	0.697
w3	0.109	0.78	0.437	0.277	0.98	0.329
workload	0.019	3.79	0.000	0.081	7.47	0.000
self	-0.001	-0.48	0.629	0.001	0.15	0.882
capi	-0.002	-1.41	0.160	-0.008	-2.32	0.020
vc	-0.248	-1.32	0.186	0.910	2.38	0.017
round	0.397	0.56	0.573	7.715	5.35	0.000
refuse	-0.021	-0.85	0.396	-0.036	-0.72	0.475
unusable	-0.518	-1.87	0.061	-0.624	-1.11	0.267
HH-contacts	3.190	7.57	0.000	0.019	0.22	0.824
σ_u	0.354			0.889		
σ_e	0.668			1.287		
ρ	0.220			0.323		
R^2	0.090			0.314		
N	305			305		
interviewer	136			136		
Box-Cox λ	0.126			0.321		
Breusch/Pagan						
χ^2_1	11.91	$(p = 0.001)$		21.94	$(p = 0.000)$	
Hausman						
χ^2_9	7.24	$(p = 0.703)$		10.23	$(p = 0.421)$	

Source: SOEP, sample E, wave 1-3 (own calculation)

Table 14: *Regression models for the transformed chi-square values of clusters in sample F, wave 1-3*

	first digit			*first two digit*		
	Random Effects Model			Fixed Effects Model		
variable	Coef.	z	$P > z$	Coef.	z	$P > z$
intercept	2.262	15.59	0.000	2.459	73.53	0.000
w2	0.006	0.17	0.868	0.000	0.01	0.989
w3	0.362	10.26	0.000	0.115	17.96	0.000
workload	0.015	14.36	0.000	0.003	5.46	0.000
self	0.001	0.80	0.422	0.000	-0.77	0.440
capi	0.000	-0.38	0.705	0.000	0.81	0.420
vc	-0.138	-2.02	0.044	0.035	2.15	0.032
round	0.794	3.17	0.002	0.573	10.84	0.000
refuse	0.006	0.81	0.415	0.001	0.54	0.590
unusable	-0.066	-0.38	0.705	-0.042	-1.31	0.189
contacts	-0.018	-1.31	0.189	-0.006	-1.77	0.078
σ_u	0.281			0.068		
σ_e	0.482			0.071		
ρ	0.254			0.478		
R^2	0.250			0.530		
N	1254			1254		
interviewer	523			523		
Boxcox λ	-			-0.273		
Breusch/Pagan						
χ^2_1	61.48	$(p = 0.000)$		43.54	$(p = 0.000)$	
Hausman						
χ^2_9	8.78	$(p = 0.553)$		18.93	$(p = 0.041)$	

Source: SOEP, sample F, wave 1-3 (own calculation)

7 Summary and conclusion

This paper focuses on fabricated interviews in the German Socio-Economic Panel (SOEP) and the detection of these falsifications. A total of 90 falsified household interviews and 184 falsified individual interviews have been detected, almost all of them in the first wave of a subsample. The share of fabricated data is low in all samples and the maximum is 2.4% in sample E. It is important to note that, apart from the fakes in sample E, falsified data have never been disseminated as part of the widely-used SOEP, since the fabrications were detected before the data were released. However, these falsifications are in the original data files - kept at DIW Berlin - and provide a rich source for methodological research.

First, we examined in detail whether Benford's Law holds in each interviewer cluster of samples A/B, C, E, and F. We find a solution to assess the plausibility of the obtained chi-square values that is independent of the cluster size. A resampling method such as the interval percentile method allows us to determine the probability $P(perc)$ of obtaining a chi-square value more extreme than that which is actually observed. High probabilities can be interpreted as a high plausibility and vice versa.

Our results show that, in fact, the fabricated clusters in samples A/B, E, and F have mostly low probabilities and occur at the top of the interviewer ranking list for the first digit distribution. If we regard the first ten interviewers as suspicious, we identify, using Benford in sample A, one of three falsifiers, in sample E wave 1, three of five falsifiers, in wave 2 one of one, and in sample F also one of one falsifiers. This is an impressive outcome.

The undetected fabricated clusters in sample E are too small for our detection procedure. However, in sample A/B, we could not find two large fabricated clusters because the first digits of their continuous values tend to conform to the logarithmic distribution.[25] However, if we relax the assumption that Benford's Law holds in the whole data set and, instead, use a more general test statistic, we find an additional falsifier among the top twenty. This test statistic only makes the assumption that the vast majority of interviewers are honest.

Finally, the most striking result is that, using Benford, we find a new fabricator who has never been detected previously by the fieldwork organization. The interviews from this cheating interviewer will be deleted in the upcoming waves of the SOEP. This success demonstrates the predictive power of our Benford method.

In the last section, we estimated linear random-effects and fixed-effects models to explain the obtained values of the chi-square statistic for both digit distributions. Our results show that several factors contribute to the values of the test statistic.

First of all, we have shown that the homogeneity in clusters is one important factor. Interviewers whose questionnaires often contain equal values or values with equal first significant digits obtain unavoidably higher chi-square values. This is important as it seems that the detection of fabrications using Benford's Law is based, among other things, on the homogeneity of the clusters.

Second, rounding of continuous values increases the chi-square values only for the first two digit distribution. It is, by far, the strongest predictor. In the last section, we showed that, for sample A/B, the first two digit distribution works better in fraud detection than the first digit distribution.[26]

Table 15 on page 47 shows the mean values of the variation and rounding coefficient for the faked and assumed non-faked clusters of samples A/B, E, and F. In all samples the mean of the variation coefficient vc in the faked clusters is lower than in the non-faked clusters. Particularly in sample F, the value of vc is only 0.054. This means that - based on the income values - the clusters of cheating interviewers are more homogeneous than non-fabricated clusters. The mean for the rounding coefficient in samples A/B and E is higher and in sample F lower in the case of

[25]We can, however, use alternatives: the examination of the plausibility for the first two digit fit statistic yields low probabilities for all three falsified clusters.

[26]Nevertheless, in samples E and F, the first digit distribution was clearly more successful than the first two digit distribution.

fabricated clusters than of the non-faked clusters.

Table 15: *Mean values of homogeneity, rounding, missing values and contacts*

	Sample A/B		Sample E		Sample F	
	non-fake	fake	non-fake	fake	non-fake	fake
vc	0.513	0.492	0.522	0.345	0.583	0.054
round	0.465	0.512	0.468	0.554	0.480	0.399
missing values	4.395	1.404	5.792	5.246	2.749	0.273
contacts	3.346	2.751	3.009	2.231	3.256	1.625

Source: SOEP, samples A/B, E, and F, wave 1 (own calculation)

Also note that the mean for missing values and household contacts are always lower in faked than in non-faked clusters. Cheating interviewers consistently underestimate these variables. Unfortunately, we cannot use this information in our Benford analysis because missing values are indicated by assigned numbers (-1;-2;-3). However, an alternative method which we called the *variability method* can also take non-continuous variables into account. The variability method is an unsupervised learning method for outlier detection. It is based on the assumption that the variability across questionnaires in faked interviews is lower than expected, considering the whole survey. The success of this method is documented in Schäfer/Schräpler/Müller/Wagner (2005). The results suggest using a combination of both procedures for detecting frauds in surveys.

References

Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572.

Biemer, P. and Stokes, S. (1989). The Optimal Design Quality Control Samples to Detect Interviewer Cheating. *Journal of Official Statistics*, 5(1):23–39.

Boyle, J. (1994). An Application of Fourier Series to the Most Significant Digit Problem. *American Mathematical Monthly*, 101:879–976.

Bredl, S., Winker, P., and Kötschau, K. (2008). A Statistical Approach to Detect Cheating Interviewers. Working Paper No. 39 of the Zentrum für internationale Entwicklungs- und Umweltforschung der Justus-Liebig-Universität Gießen.

Breusch, T. and Pagan, A. (1979). A simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47:1287–1294.

Cantwell, P. J., Bushery, J. M., and Biemer, P. (1992). Toward a Quality Improvement System for Field Interviewing: Putting Contant Reinterview Into Persepctive. *Proceedings of the American Statistical Association (Survey Research Methods Section)*, pages 74–83.

Crespi, L. (1945). The Cheater Problem in Polling. *Public Opinion Quarterly*, Winter:431–445.

Diekmann, A. (2002). Diagnose von Fehlerquellen und methodische Qualität in der sozialwissenschaftlichen Forschung. Manuskript 06/2002, Institut für Technikfolgenabschätzung (ITA). Wien.

Diekmann, A. (2007). Not the First Digit! Using Benfords Lawto Detect Fraudulent Scientific Data. *Journal of Applied Statistics*, 34:321–329.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Engel, H.-A. and Leuenberger, C. (2003). Benford's law for exponential random variables. *Statistics and Probability Letters*, 63(4):361–365.

Epanechnikov, V. (1969). Nonparametric estimation of a multidimensional probability density. *Teoriya Veroyatnostej i Ee Primeneniya*, 14:156–162.

Evans, F. B. (1961). On Interviewer Cheating. *Public Opinion Quarterly*, 25:126–127.

Greene, W. (2003). *Econometric Analysis.* Prentice Hall, Upper Saddle River, NJ, 5 edition.

Hamming, R. (1970). On the distribution of numbers. *Bell System Technical Journal*, 49:1609–1625.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Hill, T. P. (1995). A Statistical Derivation of the Significant-Digit Law. *Statistical Science*, 10:354–362.

Hill, T. P. (1996). The First-Digit Phenomen. *American Scientist*, 86:358–363.

Hill, T. P. (1999). The Difficulty of Faking Data. *Chance*, 26:8–13.

Hood, C. C. and Bushery, J. M. (1997). Getting more Bang from the Reinterview Buck: Identifying "At Risk"" Interviewers. *Proceedings of the American Statistical Association (Survey Research Methods Section)*, pages 820–824.

Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge University Press, Cambridge.

Knuth, D. (1981). *The Art of Computer Programming 2: Seminumerical Programming*. Addison-Wesly, Reading, MA.

Koch, A. (1995). Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994. *ZUMA Nachrichten*, 36:89–105.

Lolbert, T. (2008). On the non-existence of a general Benford's law. *Mathematical Social Sciences*, 55(2):103–106.

Luque, B. and Lacasa, L. (2009). The first-digit frequencies of prime numbers and Riemann zeta zeros. *Proceedings of the Royal Society A*, 465:2197–2216.

Miller, S. J. and Nigrini, M. J. (2008). The Modulu 1 Central Limit Theorem and Benford's Law for Products. *International Journal of Algebra*, 2(3):119–130.

Mochty, L. (2002). Die Aufdeckung von Manipulationen im Rechnungswesen - Was leistet das Benford's Law. *Die Wirtschaftsprüfung*, 55(14):725–736.

Moore, J. C. and Marquis, K. (1996). The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation. Working-Paper No. 212, U.S. Department of Commerce, Bureau of the Census.

Newcomb, S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics*, 4:79–83.

Nigrini, M. (1999). I've got your number. *Journal of Accountancy*, 187:79–83.

Nigrini, M. (2000). *Digital Analysis Using Benford's Law: Test & Statistics for Auditors*. Global Audit Publications, Vancouver.

Pinkham, R. (1961). On the distribution of the first significant digits. *The Annals of Mathematical Statistics*, 32:1223–1230.

Posch, P. N. (2003). Ziffernanalyse in der Fälschungsaufspürung. Das Benford-Phänomen und Steuererklärungen in Theorie und Praxis. Arbeitspapier vom Okt. 2003, Abteilung Finanzwirtschaft, University Ulm.

Reuband, K.-H. (1990). Interviews, die keine sind - "Erfolge" und "Mißerfolge" beim Fälschen von Interviews. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 4:706–733.

Rohwer, G. and Pötter, U. (2005). TDA User's Manual. Ruhr-Universität Bochum. Fakultät für Sozialwissenschaften.

Schäfer, C., Schräpler, J.-P., Müller, K.-R., and Wagner, G. G. (2005). Identification, Characteristics and Impact of Faked and Fraudulent Interviews in Surveys. *Schmollers Jahrbuch - Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 1/2005.

Schnell, R. (1991). Der Einfluß gefälschter Interviews auf Survey-Ergebnisse. *Zeitschrift für Soziologie*, 20(1):25–35.

Schräpler, J.-P. (1999). *Das Befragtenverhalten im Sozio-oekonomischen Panel: Analysen an ausgewählten Beispielen*. PhD thesis, Ruhr-Universität Bochum, Bochum.

Schräpler, J.-P. (2004a). Non-Sampling Errors in Large Panel Surveys. Income-Nonresponse, Mode effects and Fabricated Data - the case of the German Socio-Economic Panel (SOEP) and the British Household Panel Study (BHPS). Habilitation. Ruhr-University Bochum.

Schräpler, J.-P. (2004b). Respondent behavior in Panel Studies - A Case Study for Income-Nonresponse by Means of the German Socio-Economic Panel (SOEP). *Sociological Methods & Research.*, 33(1):118–156.

Schräpler, J.-P. and Wagner, G. G. (2001). Das "Interviewer-Panel" des Sozio-oekonomischen Panels - Darstellung und ausgewählte Analysen. *Allgemeines Statistisches Archiv*, 85(1).

Schräpler, J.-P. and Wagner, G. G. (2005). Characteristics and Impact of Faked Interviews in Surveys - An analysis of genuine fakes in the raw data of SOEP. *Allgemeines Statistisches Archiv*, 89(1):7–20.

Schreiner, I., Pennie, K., and Newbrough, J. (1988). Interviewer faslification in Census Bureau Surveys. *Proceedings of the American Statistical Association (Survey Research Methods Section)*, pages 491–496.

Scott, P. and Fasli, M. (2001). Benford's Law: An Empirical Investigation and a Novel Explanation. CSM Technical Report 349, Department of Computer Science, University Essex.

StataCorp (2003). *Stata Base Reference Manual*. Stata Press, Lakeway Drive, Texas, USA.

Stokes, L. S. and Jones, P. (1989). Evaluation of the Interviewer Quality Control Procedure for the Post-Enumeration Survey. *Proceedings of the American Statistical Association (Survey Research Methods Section)*, pages 696–198.

Swanson, D., Cho, M. J., and Eltinge, J. (2003). Detecting possibly fraudulent or error-prone survey data using Benford's Law. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Tödter, K.-H. (2009). Benford's Law as an Indicator of Fraud in Economics. *German Economic Review*, 10(3):339–351.

Turner, C. F., Gribble, J. N., Al-Tayyib, A. A., and Chromy, J. R. (2002). Falsification in Epidemiologic Surveys: Detection and Remediation (Prepublication Draft). Technical Papers on Health and Behavior Measurement, No. 53. Washington DC: Research Triangle Institute.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements. *Schmollers Jahrbuch. Journal of Applied Social Science Studies*, 127(1):161–191.

A Appendix

List of Tables

Table 16: *The joint distribution for the first two digits in according with Benford*

D_1	D_2: 0	1	2	3	4	5	6	7	8	9	$\sum_{D_2=0}^{D_2=9}$
1	0.0413927	0.0377886	0.0347621	0.0321847	0.0299632	0.0280287	0.0263289	0.0248236	0.0234811	0.0222764	0.3010300
2	0.0211893	0.0202034	0.0193052	0.0184834	0.0177288	0.0170333	0.0163904	0.0157943	0.0152400	0.0147233	0.1760913
3	0.0142404	0.0137883	0.0133640	0.0129650	0.0125891	0.0122345	0.0118992	0.0115819	0.0112810	0.0109954	0.1249387
4	0.0107239	0.0104654	0.0102192	0.0099842	0.0097598	0.0095453	0.0093400	0.0091434	0.0089548	0.0087739	0.0969100
5	0.0086002	0.0084332	0.0082725	0.0081179	0.0079689	0.0078253	0.0076868	0.0075531	0.0074240	0.0072992	0.0791812
6	0.0071786	0.0070619	0.0069489	0.0068394	0.0067334	0.0066306	0.0065309	0.0064341	0.0063402	0.0062489	0.0669468
7	0.0061603	0.0060741	0.0059904	0.0059089	0.0058295	0.0057523	0.0056771	0.0056039	0.0055325	0.0054629	0.0579919
8	0.0053950	0.0053288	0.0052642	0.0052012	0.0051396	0.0050795	0.0050208	0.0049634	0.0049073	0.0048525	0.0511525
9	0.0047989	0.0047464	0.0046951	0.0046449	0.0045958	0.0045476	0.0045005	0.0044543	0.0044091	0.0043648	0.0457575
$\sum_{D_1=1}^{D_1=9}$	0.1196793	0.1138901	0.1088215	0.1043296	0.1003082	0.0966772	0.0933747	0.0903520	0.0875701	0.0849974	

Source: own calculation

Marginal distribution for first to fourth significant digit d

d	$P(D_1 = d)$	$P(D_2 = d)$	$P(D_3 = d)$	$P(D_4 = d)$	uniform distribution
0	-	0.1196793	0.1017843	0.1001761	0.1000000
1	0.3010300	0.1138901	0.1013759	0.1001368	0.1000000
2	0.1760913	0.1088215	0.1009721	0.1000976	0.1000000
3	0.1249387	0.1043296	0.1005729	0.1000584	0.1000000
4	0.0969100	0.1003082	0.1001780	0.1000193	0.1000000
5	0.0791812	0.0966772	0.0997870	0.0999802	0.1000000
6	0.0669468	0.0933747	0.0994013	0.0999412	0.1000000
7	0.0579919	0.0903520	0.0990192	0.0999022	0.1000000
8	0.0511525	0.0875701	0.0986411	0.0998632	0.1000000
9	0.0457575	0.0849974	0.0982671	0.0998243	0.1000000

Source: own calculation

Table 17: *Descriptive measurements of HH-contacts in each subsample*

	Obs.	Mean	Std. Dev.	Min	Max
Sample A/B					
wave 1	631	3.342	1.328	1	13
wave 2	451	2.675	0.934	1	7
wave 3	397	2.681	1.171	1	9
Sample C					
wave 1	214	n.k.	n.k.	n.k.	n.k.
wave 2	215	2.199	0.711	1	5.25
wave 3	270	2.473	0.996	1	9
Sample E					
wave 1	128	2.997	1.168	1.1	8.8
wave 2	131	3.145	1.424	1	9
wave 3	132	2.776	1.172	1	7
Sample F					
wave 1	539	3.253	1.289	1	9
wave 2	493	3.012	1.289	1	9
wave 3	469	2.796	1.197	1	8

Source: SOEP samples A/B, C, E, and F, (own calcul.)

Table 18: *Decriptive measurements of the average number of refused answers in clusters*

	Obs.	Mean	Std. Dev.	Min	Max
Sample A/B					
wave 1	636	4.377	1.929	0	23.00
wave 2	463	1.903	1.991	0	18.00
wave 3	410	1.828	1.974	0	30.00
Sample C					
wave 1	214	2.279	2.800	0	20.22
wave 2	264	1.567	1.598	0	13.50
wave 3	278	2.003	2.021	0	13.55
Sample E					
wave 1	150	5.774	3.178	2	19.00
wave 2	125	2.255	1.783	0	8.62
wave 3	129	1.864	1.698	0	8.43
Sample F					
wave 1	536	2.744	2.367	0	25.00
wave 2	473	2.894	2.727	0	18.00
wave 3	461	2.760	2.381	0	21.86

Source: SOEP samples A/B, C, E, and F, indiv. questionnaires (own calc.)

Table 19: *Selected continuous variables for the Benford analysis*

Sample A/B, wave 1		Sample E, wave 1		Sample E, wave 2	
Variable	Label	Variable	Label	Variable	Label
ap3301	gross income last month	op1801	monthly salary at which would take job	pp1801	monthly salary at which would take job
ap3302	net income last month	op4501	gross income last month	pp6001	gross income last month
ap3904	gross amount 14th month pay, prev. Yr.	op4502	net income last month	pp6002	net income last month
ap3906	gross amount of christmas bonus, prev. Yr.	op4902	gross amount second job monthly income	pp6602	gross amount second job monthly income
ap3908	gross amount of vacation bonus, prev. Yr.	op4904	gross amount of unemployment benefit, monthly	pp6604	gross amount of unemployment benefit, monthly
ap3110	gross amount special allowance	op4906	gross amount of unemployment relief, monthly	pp6606	gross amount of unemployment relief, monthly
ap4307	amount of health insurance premium	op4908	gross amount of subsistence allowance,monthly	pp6608	gross amount of subsistence allowance,monthly
ap4405	amount of voluntary, Compulsory Premium	op4910	gross amount early retirement benefit, monthly	pp6610	gross amount of transition money etc.
ap46a02	amount of taxes	op4912	gross amount old-age, invalid pension, monthly	pp6612	gross amount early retirement benefit, monthly
ap46b02	amount jointly assessed	op4914	gross emount widow-er, orphan benefit, monthly	pp6614	gross amount old-age, invalid pension, monthly
ap46b03	amount jointly to be refunded	op4916	gross amount maternity benefit	pp6616	gross emount widow-er, orphan benefit, monthly
ap46b04	amount assessed single	op4918	gross amount student grant	pp6618	gross amount maternity benefit, prev. Year
ap46b05	amount to be refunded single	op4920	gross amount military Comm. Serv. Pay	pp6620	gross amount student grant, prev. Yr.
ap4702	amount of payments to parents, prv. Year	op4922	gross amount income-persons not in HH., monthly	pp6622	gross amount military, comm. serv. pay, prev. Yr.
ap4704	amount of payments to children, prev. Yr.	op5602	amount of settlement from last company	pp6624	gross amount income person not in HH., prev. Yr.
ap4706	amount of payments to ex-spouse, prev. Yr.	op58a03	gross amount of wages, salary, prev. Yr.	pp7302	amount of severance package, compensation
ap4708	amount of payments to relatives, prev. Yr.	op58b03	gross amount of self-employed income, prev. Yr.	pp76a03	gross amount of wages, salary, prev. Yr.
ap4710	amount of payments to others, prev. Yr.	op58c03	gross amount second job income, prev. Yr.	pp76b03	gross amount of self-employed income, prev. Yr.
ap47a02	amount send to native country, prev. Yr.	op58d03	gross amount old age, invalid pension, prev. Yr.	pp76c03	gross amount second job income, prev. Yr.
ap47a03	amount of money given family, prev. Yr.	op58e03	gross amount widow-er. orphan benefit, prev. Yr.	pp76d03	gross amount old age, invalid pension, prev. Yr.
ap47a04	amount of money towards saving, prev. Yr.	op58f03	gross amount unemployed benefit, prev. Yr.	pp76e03	gross amount widow-er. orphan benefit, prev. Yr.
ap47a05	amount of money given to others, prev. Yr.	op58g03	gross amount unemployed relief, prev. Yr.	pp76f03	gross amount unemployed benefit, prev. Yr.
ap47b03	amount to German relatives not in HH.	op58h03	gross amount maint. support, trnng., prev. Yr.	pp76g03	gross amount unemployed relief, prev. Yr.
		op58i03	gross amount early retirement benefit, prev. Yr	pp76h03	gross amount maint. support, trnng., prev. Yr.
		op58j03	gross amount maternity benefit, prev. Year	pp76i03	gross amount early retirement benefit, prev. Yr
		op58k03	gross amount student grant, prev. Yr.	pp76j03	gross amount maternity benefit, prev. Year
		op58l03	gross amount military, comm. serv. pay, prev. Yr.	pp76k03	gross amount student grant, prev. Yr.
		op58m03	gross amount income person not in HH., prev. Yr.	pp76l03	gross amount military, comm. serv. pay, prev. Yr.
		op5902	amount of 13th month pay, prev. Yr.	pp76m03	gross amount income person not in HH., prev. Yr.
		op5904	amount of 14th. Month pay, prev. Yr.	pp7702	amount of 13th month pay, prev. Yr.
		op5906	amount of christmas bonus, prev. Yr.	pp7704	amount of 14th. Month pay, prev. Yr.
		op5908	amount of vacation bonus, prev, Yr.	pp7706	amount of christmas bonus, prev. Yr.
		op5910	amount of profit-sharing bonus, prev. Yr.	pp7708	amount of vacation bonus, prev, Yr.
		op5912	amount of other bonuses, prev. Yr.	pp7710	amount of profit-sharing bonus, prev. Yr.
		op6101	amount of worker pension-self	pp7712	amount of other bonuses, prev. Yr.
		op6102	amount of miner pension-self	pp7901	amount of worker pension-self
		op6103	amount of civil servant pension-self	pp7902	amount of miner pension-self
		op6104	amount of war viction pension-self	pp7903	amount of civil servant pension-self
		op6105	amount of farmer pension-self	pp7904	amount of war viction pension-self
		op6106	amount of worker accident pension-self	pp7905	amount of farmer pension-self
		op6107	amount of supp. Civil servant pension-self	pp7906	amount of worker accident pension-self
		op6108	amount of company pension-self	pp7907	amount of supp. Civil servant pension-self
		op6109	amount of other pension-self	pp7908	amount of company pension-self
		op6110	amount of worker pension-dependent	pp7909	amount of other pension-self
		op6111	amount of miner pension-dependent	pp7910	amount of worker pension-dependent
		op6112	amount of civil servant pension-dependent	pp7911	amount of miner pension-dependent
		op6113	amount of war-victim pension-dependent	pp7912	amount of civil servant pension-dependent
		op6114	amount of farmer pension-dependent	pp7913	amount of war-victim pension-dependent
		op6115	amount of worker accident pension-dependent	pp7914	amount of farmer pension-dependent
		op6116	amount of supplementary pension-dependent	pp7915	amount of worker accident pension-dependent
		op6117	amount of company pension-dependent	pp7916	amount of supplementary pension-dependent
		op6118	amount of other pension-dependent	pp7917	amount of company pension-dependent
		op6309	amount of monthly health insurance premium	pp7918	amount of other pension-dependent
		op6312	amount retained	pp8401	amount additional private insurance
		op12206	amount of payments to children, prev. Yr.	pp9001	amount private health insurance
		op12210	amount of payments to spouse, prev. Yr.	pp13402	amount of payments to parents, prev. Yr.
		op12218	amount of payments to others, prev. Yr.	pp13406	amount of payments to children, prev. Yr.
				pp13410	amount of payments to spouse, prev. Yr.
				pp13414	amount of payments to relatives, prev. Yr.
				pp13418	amount of payments to others, prev. Yr.

Source: SOEP, samples A/B and E, individual questionnaire

Table 20: *Interviewer ranking by plausibility, first two digit distribution, sample A/B, wave 1*

Rank	Intnr	digits	chi-sq.	P(normal)	P(perc)	chi-sq. (boot)
1	xx365x	227	831.9497	1.83E-11	0	410.2177
2	xx582x	170	709.7847	2.22E-16	0	280.2723
3	xx049x	77	402.2741	1.05E-06	0	208.797
4	xx143x	203	626.5366	4.30E-07	0	358.1882
5	xx512x	90	418.3447	5.33E-07	0	216.8385
6	xx552x	3	267.5491	0.000138	0	93.35402
7	**xx827x**	**122**	**453.0379**	**2.52E-06**	**0.001**	**252.0209**
8	xx901x	146	445.6096	9.61E-05	0.001	281.5031
9	xx698x	63	388.9238	0.000118	0.004	220.3225
10	xx202x	213	557.7646	0.000356	0.005	366.5032
11	**xx582x**	**152**	**416.7733**	**0.00108**	**0.005**	**285.5435**
12	**xx800x**	**95**	**371.1279**	**0.000179**	**0.006**	**221.9303**
13	xx147x	94	352.1639	0.001545	0.009	224.0793
14	xx756x	58	390.9862	0.00094	0.009	230.8448
15	xx079x	139	382.5185	0.003442	0.01	265.6753
16	xx474x	147	411.88	0.001418	0.012	283.9175
17	xx306x	104	323.4291	0.003711	0.016	218.1264
18	xx801x	70	314.7503	0.007359	0.02	213.2567
19	xx515x	78	301.8357	0.012839	0.026	210.5901
20	xx852x	63	320.2045	0.014672	0.029	220.3225
21	xx401x	26	328.9332	0.010712	0.029	197.208
22	xx544x	76	300.4866	0.016797	0.031	211.4826
23	xx320x	16	301.4572	0.020058	0.043	177.4095
24	xx709x	65	312.2405	0.029421	0.043	225.6979
25	xx895x	118	319.3288	0.031588	0.044	241.3161
26	xx234x	55	311.488	0.040255	0.049	222.2785
27	xx815x	69	287.4532	0.042703	0.053	215.3047
28	xx293x	101	286.6554	0.040163	0.059	218.1927
29	xx450x	96	289.7921	0.047038	0.063	221.3673
30	xx112x	45	303.869	0.054023	0.064	223.1108
			⋮			
636	xx087x	46	87.73752	0.996945	1	222.4255

Source SOEP, individual questionnaire, only continuous variables, 1984 (own calcul.)

Table 21: *Interviewer ranking by Plausibility, first two-digit distribution, sample C, wave 1*

Rank	Intnr	digits	chi-sq.	P(normal)	P(perc)	chi-sq. (boot)
1	xx053x	115	664.9291	2.22E-16	0	284.3378
2	xx611x	99	824.206	2.22E-16	0	271.0816
3	xx528x	107	639.6402	2.22E-16	0	275.69
4	xx303x	98	510.0764	5.63E-08	0.0001	266.915
5	xx246x	58	478.6509	2.42E-07	0.0008	231.4268
6	xx056x	67	460.2145	3.23E-05	0.0029	253.7367
7	xx211x	91	412.443	0.000113	0.0036	249.3335
8	xx076x	80	391.9182	0.000788	0.0078	247.7214
9	xx670x	57	377.2998	0.002031	0.0124	233.8839
10	xx840x	87	381.5224	0.003785	0.015	255.8684
11	xx622x	94	372.2561	0.004225	0.0155	253.5131
12	xx567x	64	393.5778	0.004322	0.017	253.3455
13	xx617x	76	373.1486	0.005858	0.0191	250.6293
14	xx408x	9	195.0309	0.003692	0.0209	104.216
15	xx202x	43	301.8178	0.011837	0.0253	213.0471
16	xx452x	69	350.4006	0.019552	0.0381	245.2613
17	xx571x	88	341.8425	0.028047	0.0451	253.8837
18	xx031x	96	351.6598	0.036524	0.0498	267.2451
19	xx443x	46	274.9916	0.034572	0.0501	206.8606
20	xx606x	44	271.4332	0.048766	0.0624	206.904
21	xx311x	71	328.9495	0.066986	0.0796	254.1529
22	xx811x	92	312.5017	0.073781	0.081	248.8333
23	xx151x	48	249.0779	0.086152	0.0948	198.3651
24	xx264x	91	296.061	0.145357	0.1402	249.3335
25	xx431x	42	252.601	0.161368	0.1575	213.7435
26	xx142x	103	310.4689	0.204754	0.1879	272.8453
27	xx111x	51	245.8508	0.204872	0.1913	212.8777
28	xx180x	55	276.2341	0.222694	0.192	235.6924
29	xx754x	59	277.3699	0.227564	0.1975	239.0226
30	xx959x	79	279.7506	0.24075	0.2057	246.8276
			⋮			
214	xx377x	125	131.1992	0.999769	1	298.4655

Source SOEP, sample C, wave1, only cont. variables, 1989 (own calcul.)

Table 22: *Interviewer ranking by Plausibility, first two digit distribution, sample F, wave 1*

Rank	Intnr	digits	chi-sq.	P(normal)	P(perc)	chi-sq. (boot)
1	xx690x	124	638.5831	7.23E-09	0	334.9819
2	xx199x	3	553.0672	2.62E-05	0	222.4406
3	xx708x	106	569.7923	3.70E-08	0	301.6514
4	xx057x	75	547.3914	2.94E-06	0.0005	306.9953
5	xx129x	98	529.2237	3.78E-06	0.001	303.1539
6	xx314x	95	523.2931	8.48E-06	0.001	303.1845
7	xx984x	9	603.0896	0.000375	0.0095	251.1293
8	xx027x	24	511.817	0.002129	0.011	287.4346
9	xx232x	140	516.08	0.005096	0.0145	369.8288
10	xx127x	124	473.4838	0.00486	0.016	334.9819
11	xx739x	90	399.9429	0.015306	0.03	292.4988
12	xx046x	53	368.1531	0.024254	0.0405	256.8608
13	xx451x	78	394.6435	0.027093	0.0425	299.7499
14	xx382x	78	390.4618	0.032848	0.0465	299.7499
15	xx951x	85	390.8057	0.034371	0.048	297.6858
16	xx076x	65	371.7019	0.038472	0.0515	282.051
17	xx564x	120	416.4118	0.039299	0.053	324.2004
18	xx476x	91	377.1788	0.037939	0.054	291.022
19	xx262x	71	372.6702	0.055355	0.066	292.5239
20	xx679x	23	383.3964	0.075064	0.0875	269.0439
21	xx911x	55	342.3433	0.086705	0.089	265.8463
22	xx079x	136	440.1743	0.084204	0.0915	361.5663
23	xx393x	115	364.206	0.094594	0.0965	300.63
24	xx257x	33	365.9924	0.095279	0.1015	277.0514
25	xx315x	59	351.1641	0.094835	0.1025	277.9259
26	xx063x	133	427.5619	0.098681	0.103	351.5313
27	xx236x	19	311.8948	0.108755	0.1095	226.4403
28	xx668x	163	462.8384	0.123697	0.1245	396.1332
29	xx800x	183	527.7181	0.134018	0.1375	453.1797
30	xx249x	105	345.0767	0.157496	0.147	295.9721
31	xx404x	20	308.9099	0.178108	0.1655	242.8545
32	xx831x	74	349.1529	0.188435	0.1795	302.8532
33	xx802x	31	327.4075	0.2379	0.2075	275.5537
34	xx009x	54	309.0565	0.251996	0.23	270.0162
35	xx213x	74	334.7756	0.271172	0.238	302.8532
36	xx115x	92	336.8014	0.259955	0.2385	303.1825
37	xx864x	10	299.9186	0.296701	0.2495	245.7578
38	xx136x	161	427.516	0.27826	0.2565	394.8668
39	xx924x	14	275.7848	0.298095	0.257	233.1723
40	xx787x	45	296.9131	0.283414	0.261	263.5864
41	xx986x	14	271.17	0.318284	0.274	233.1723
42	xx029x	267	596.3646	0.293913	0.2755	554.0989
43	xx660x	38	292.6412	0.335834	0.3025	266.8111
44	xx815x	84	320.5421	0.341419	0.3095	298.9255
45	xx796x	102	316.4268	0.347767	0.314	296.5205
46	xx839x	88	304.8939	0.35734	0.3175	286.8169
47	xx389x	87	308.5076	0.358326	0.3225	290.3978
48	xx062x	122	351.1829	0.361406	0.3255	331.6586
49	xx609x	86	310.6596	0.377863	0.3375	295.0213
50	xx876x	42	294.1547	0.403438	0.3495	278.8361
51	xx843x	114	310.3228	0.394953	0.35	297.5998
52	xx552x	44	284.5226	0.409528	0.353	271.0785
53	xx617x	53	269.7471	0.409653	0.355	256.8608
54	xx010x	124	349.6234	0.392297	0.3575	334.9819
55	xx046x	6	314.8121	0.392405	0.366	284.9588
56	xx977x	54	283.2419	0.410456	0.3665	270.0162
57	xx282x	139	384.9927	0.399867	0.367	370.5132
58	**xx085x**	**43**	**284.0233**	**0.436945**	**0.383**	**274.4162**
59	xx436x	64	288.9794	0.429788	0.3875	279.6033
60	xx663x	20	250.3141	0.45851	0.3885	242.8545
61	xx524x	46	268.2891	0.451171	0.403	261.0893
62	xx491x	78	305.261	0.455484	0.408	299.7499
63	xx044x	14	235.984	0.486054	0.4085	233.1723
64	xx343x	114	303.9369	0.447213	0.4095	297.5998
65	xx496x	48	260.2462	0.459289	0.41	254.6699
66	xx273x	22	264.5579	0.47853	0.4145	260.2214
67	xx242x	224	524.5763	0.46759	0.4155	518.1083
68	xx995x	138	372.1559	0.45608	0.4155	365.5529
69	xx278x	63	283.9802	0.465071	0.4265	279.3771
70	xx642x	292	616.3593	0.463819	0.4265	608.1957
71	xx724x	99	303.1847	0.477636	0.437	300.3804
72	xx037x	50	263.8566	0.491234	0.4395	262.6082
73	xx491x	40	278.3183	0.520932	0.4695	281.6244
74	xx635x	86	292.7491	0.518027	0.4795	295.0213
75	xx251x	46	255.1027	0.540633	0.493	261.0893
76	xx325x	60	270.0982	0.542965	0.4975	276.0948
77	xx454x	112	288.0436	0.550965	0.5125	294.2781
78	xx589x	13	221.7654	0.597106	0.514	242.9841
79	xx530x	36	263.7028	0.567788	0.515	274.8634
80	xx969x	73	287.0598	0.557856	0.5215	294.5273
			...			
536	xx260x	29	119.1835	0.979552	1	279.8405

Source SOEP, individual questionnaire, only continuous variables, 2000 (own calcul.)